Steady-State Simulation of Queueing Processes: A Survey of Problems and Solutions

KRZYSZTOF PAWLIKOWSKI

Department of Computer Science, University of Canterbury, Christchurch, New Zealand

For years computer-based stochastic simulation has been a commonly used tool in the performance evaluation of various systems. Unfortunately, the results of simulation studies quite often have little credibility, since they are presented without regard to their random nature and the need for proper statistical analysis of simulation output data.

This paper discusses the main factors that can affect the accuracy of stochastic simulations designed to give insight into the steady-state behavior of queuing processes. The problems of correctly starting and stopping such simulation experiments to obtain the required statistical accuracy of the results are addressed. In this survey of possible solutions, the emphasis is put on possible applications in the sequential analysis of output data, which adaptively decides about continuing a simulation experiment until the required accuracy of results is reached. A suitable solution for deciding upon the starting point of a steady-state analysis and two techniques for obtaining the final simulation results to a required level of accuracy are presented, together with pseudocode implementations.

Categories and Subject Descriptors: G.3 [**Probability and Statistics**]: Statistical Computing, Statistical Software; G.4 [**Mathematical Software**]: Efficiency, Reliability and Robustness; G.m [**Miscellaneous**]: Queuing Theory; I.6.4 [**Simulation and Modeling**]: Model Validation and Analysis

General Terms: Algorithms, Performance, Theory

Additional Key Words and Phrases: Automation of simulation experiments, initial transient period, precision of simulation results, sequential analysis of confidence intervals, statistical analysis of simulation output data, stopping rules for steady-state simulation

INTRODUCTION

Computer-based stochastic simulation, traditionally regarded as a last resort tool (if analytical methods fail), has become a valid and commonly used method of performance evaluation. This popularity is due to the continuing development of more powerful and less expensive computers, as well as significant achievements in software engineering. One can observe a trend toward integrating simulation methodology with concepts and methods of artificial intelligence [Artificial Intelligence 1988]. Various user-friendly simulation packages offer visual interactive capabilities; traditional discrete-event simulation modeling is more and more frequently supported by object and logic-oriented programming and various concepts of artificial intelligence [Bell and O'Kneefe 1987; Gates et al. 1988; Jackman and Medeiros 1988; Kerckhoffs and Vansteenkiste 1986; Knapp 1986; Ören and Zeigler 1987; Reedy 1987; Ruiz-Mier

CONTENTS

and Talavage 1987; Stairmand and Kreutzer 1988; Zeigler 1987]. All of these developments offer users increasingly powerful and versatile techniques for performance evaluation, leading toward automatic, knowledge-based simulation packages. Simulation programming techniques and languages are discussed in numerous publications including textbooks by Bulgren [1982], Kreutzer [1986], Law and Kelton [1982a], and Payne [1982].

Applying simulation to the modeling and performance analysis of complex systems can be compared to using the surgical scalpel [Shannon 1981], whereby "in the right hand [it] can accomplish tremendous good, but it must be used with great care and by someone who knows what they are doing." One of the applications in which simulation has become increasingly popular is the class of dynamic systems with random input and output processes, represented for example by computer communication networks. In such cases, regardless of how advanced the programming methodology applied to simulation modeling is, since simulated events are controlled by random numbers, the results produced are nothing more than statistical samples. Therefore, various simulation studies, frequently reported in technical literature, can be regarded as programming exercises only. The authors of such studies, after putting much intellectual effort and time into building simulation models and then writing and running programs, have very little or no interest in a proper analysis of the simulation results. It is true that "the purpose of modeling is insight, not numbers" [Hamming 1962], but proper insight can only be obtained from correctly analyzed numbers. Other modes of presenting results, for example, animation, can be very attractive and useful when the model is validated, but nothing can substitute the need for statistical analysis of simulation output data in studies aimed at performance analysis; see also Schruben [1987].

In the stochastic simulation of, for example, queuing systems "computer runs yield a mass of data but this mass may turn into a mess." If the random nature of the results is ignored, "instead of an expensive simulation model, a toss of the coin had better be used" [Kleijnen, 1979]. Statistical inference is an absolute necessity in any situation when the same (correct) program produces different (but correct) output data from each run. Any sequence $x_1, x_2, \ldots, x_n$ of such output data simply consists of realizations of random variables $X_1, X_2, \ldots, X_n$. Examples illustrating this fact may be found in Kelton [1986], Law [1983], Law and Kelton [1982a], and Welch [1983, Sec. 6.1].

The simplest objective of simulation studies is the estimation of the mean μ_x of an analyzed process from the sequence of collected observations $x_1, x_2, \ldots, x_n$, by calculating the average as follows:

$$\bar{X}(n) = \sum_{i=1}^{n} \frac{x_i}{n}. \tag{1}$$

Such an average assumes a random value, which depends on the sequence of observations. The accuracy with which it estimates an unknown parameter μ_x can be assessed by the probability

$$P(|\bar{X}(n) - \mu_x| < \Delta_x) = 1 - \alpha \tag{2a}$$

or

$$P(\bar{X}(n) - \Delta_x \le \mu_x \le \bar{X}(n) + \Delta_x) = 1 - \alpha, \tag{2b}$$

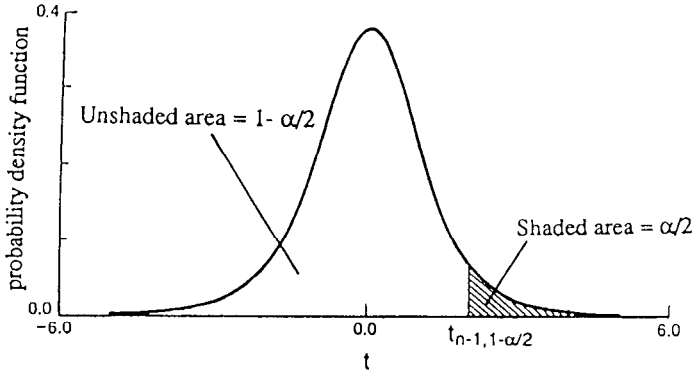

Figure 1. $P[T_{n-1} \leq t_{n-1,1-\alpha/2}] = 1 - \alpha/2$.

where Δ_x is the half-width of the confidence interval for the estimator and $(1 - \alpha)$ is the confidence level, $0 < \alpha < 1$. Thus, if the width $2\Delta_x$ of the confidence interval is found for an assumed confidence level of $(1 - \alpha)$ and the simulation experiment were repeated a number of times, the interval $(\bar{X}(n) - \Delta_x, \bar{X}(n) + \Delta_x)$ would contain an unknown average μ_x in $100(1 - \alpha)\%$ of cases and would not in $100\alpha\%$ of cases. It is well known that if observations $x_1, x_2, \ldots, x_n$ can be regarded as realizations of independent and normally distributed random variables $X_1, X_2, \ldots, X_n$, then

$$\Delta_x = t_{n-1,1-\alpha/2}\hat{\sigma}[\bar{X}(n)], \tag{3}$$

where

$$\hat{\sigma}^2[\bar{X}(n)] = \sum_{i=1}^{n} \frac{\{x_i - \bar{X}(n)\}^2}{n(n-1)} \tag{4}$$

is the (unbiased) estimator[1] of the variance of $\bar{X}(n)$, and $t_{n-1,1-\alpha/2}$ is the upper $(1 - \alpha/2)$ critical point obtained from the t-distribution with $(n - 1)$ degrees of freedom. In other words, for given $1 - \alpha/2$, assuming the t-distribution for the random variable $T_{n-1} = (\bar{X}(n) - \mu_x)/\hat{\sigma}[\bar{X}(n)]$, we get $P[T_{n-1} \leq t] = 1 - \alpha/2$ for $t = t_{n-1,1-\alpha/2}$; see Figure 1 and Appendix A. For this reason, $t_{n-1,1-\alpha/2}$ is also called the $(1 - \alpha/2)$ quantile, or percentile, of the t-distribution with $(n - 1)$ degrees of freedom. For $n > 30$, the t-distribution can be replaced by the standard normal distribution. In that case $t_{n-1,1-\alpha/2}$ in Equation (3) should be replaced by $z_{1-\alpha/2}$, which is the upper $(1 - \alpha/2)$ critical point obtained from the standard normal distribution or, equivalently, the $(1 - \alpha/2)$ quantile of the standard normal distribution; see Appendix A. Commonly used values of $t_{n-1,1-\alpha/2}$ and $z_{1-\alpha/2}$ have been tabularized and can be found in many textbooks; see, for example, Trivedi [1982, Appendix 3, and p. 489]. (Warning: The definitions used for obtaining tabularized values should always be checked. For example, $t_{n-1,1-\alpha/2}$ and $z_{1-\alpha/2}$ are sometimes denoted as $t_{n-1,\alpha/2}$ and $z_{\alpha/2}$, respectively.)

Equation (3) can also be applied if the observations $x_1, x_2, \ldots, x_n$ represent random variables that are not normally distributed. That is, if the observations are realizations of independent and identically distributed (i.i.d.) random variables $X_1, X_2, \ldots, X_n$, then according to the central limit theorem (see Appendix A), the distribution of the variable $\bar{X}(n)$ tends to the normal distribution as the number of collected observations n increases. In practice, Equation (3) gives a good approximation for $n > 100$. Results obtained from Equations (1) and (3) are called *point* and *interval estimates*, respectively. Both of them are important: The former characterizes the system analyzed, and the latter states the accuracy of the obtained characteristics.

If observations $x_1, x_2, \ldots, x_n$ cannot be regarded as realizations of i.i.d. random

[1] Following standard notation, $\hat{a}$ means an estimator of the parameter a.

variables, we have to consider some modifications to the above estimators. This raises the problem of measuring the quality of estimators. There are three common measures of estimator effectiveness:

(1) The *bias*, which measures the systematic deviation of the estimator from the true value of the estimated parameter; for example, in the case of $\bar{X}(n)$,

$$\text{Bias}[\bar{X}(n)] = \text{E}[\bar{X}(n) - \mu_x]. \quad (5)$$

(2) The *variance*, which measures the mean (squared) deviation of the estimator from its mean value; that is,

$$\sigma^2[\bar{X}(n)] = \text{E}[\{\bar{X}(n) - \text{E}[\bar{X}(n)]\}^2]. \quad (6)$$

(3) The *mean square error* (MSE) of the estimator, defined as

$$\text{MSE}[\bar{X}(n)] = \text{E}\{[\bar{X}(n) - \mu_x]^2\}. \quad (7)$$

Note that from these definitions,

$$\text{MSE}[\bar{X}(n)] = \{\text{Bias}[\bar{X}(n)]\}^2 + \sigma^2[\bar{X}(n)]. \quad (8)$$

The main analytical problem encountered in the analysis of simulation results is that they are usually highly correlated and thus do not satisfy the precondition of statistical independence. If observations $x_1, x_2, \ldots, x_n$ represent an autocorrelated and stationary sequence of random variables $X_1, X_2, \ldots, X_n$, then the variance of $\bar{X}(n)$ is given by the formula

$$\sigma^2[\bar{X}(n)] = \left[\text{R}(0) + 2 \cdot \sum_{k=1}^{n-1} \left(1 - \frac{k}{n}\right)\text{R}(k)\right] \Big/ n, \quad (9)$$

where

$$\text{R}(k) = \text{E}[(X_i - \mu_x)(X_{i-k} - \mu_x)], \quad 0 \le k \le n-1 \quad (10)$$

is the autocovariance of order k (the lag k component of the autocorrelation function $\{\text{R}(k)\}$) of the sequence. The autocovariances defined in Equation (10) are independent of the index i because of the assumed stationarity of the analyzed processes. Note that the variance $\sigma^2[\bar{X}(n)]$ can be reduced to $\text{R}(0)/n$, and consequently could be estimated by Equation (4), if and only if the observations are uncorrelated. Neglecting the existing statistical correlation is equivalent to removing all the components except R(0) from Equation (9). Such an approximation is usually unacceptable. For example, in an M/M/1 queuing system with 90% utilization, the variance of the mean queue length calculated according to Equation (9) is 367 times greater than that from Equation (4) [Blomqvist 1967]; see Law and Kelton [1982a, p. 146] for another example. Any variance analysis disregarding correlations among the observations would lead either to an excessively optimistic confidence interval for μ_x, in the case of positively correlated observations, or to an excessively pessimistic confidence interval for μ_x, in the case of negatively correlated observations; see Equation (3). A positive correlation between observations is typical in simple queuing systems without feedback connections and is stronger for higher system utilization; see, for example, Daley [1968] for correlation analysis of the M/M/1 queue.

Generally, the analysis of variance of correlated processes, and the analysis of their autocorrelation functions in particular, is a complex statistical problem and therefore creates a major difficulty in the statistical analysis of simulation output data. In *terminating* (or *finite-horizon*) *simulation* used for studying the behavior of systems during specified intervals of time, the above problem can be overcome by making a number of independent replications of the simulation experiment. In that case the means of individual observations collected during different simulation runs can be regarded as a sequence of independent (secondary) output data, and Equation (4) can be applied. Exhaustive discussions on the statistical analysis of output data from terminating simulation can be found, for example, in Kleijnen [1979, 1987], Law [1980], and Law and Kelton [1982a, Sec. 8.5].

In this paper, we discuss *steady-state* (*infinite-horizon*) *simulation*, aimed to give insight into the behavior of queuing

processes after a long period of time. The methodology for this kind of simulation study is complicated. After launching, a queuing process is initially in a nonstationary phase (warm-up period). Then, if the process is stable, it moves asymptotically toward a steady state (statistical equilibrium), although different parameters usually tend to the steady state with different rates. Since observations gathered during the initial transient periods do not characterize the steady state, a natural idea is to discard all such observations before further analysis. This requires an estimation of the effective length of the initial transient period. Ignoring the existence of this period can lead to a significant bias of the final results. On the other hand, the removal of any observations increases the variance of estimates, which in turn can increase the value of the mean-square error [Donnelly and Shannon 1981; Fishman 1972; Turnquist and Sussman 1977; Wilson and Pritsker 1978a]. Thus, a decision of whether to delete or not to delete initial observations depends on the assumed criterion of goodness of the estimators. This also affects methods used to collect observations, which are discussed in Section 1. These and other aspects of the problem of initialization are presented broadly in Section 2.

Several methods of data collection and analysis have been proposed to overcome the theoretical problems that arise from the correlated nature of observations collected during steady-state simulation. We survey these methods in Section 1. They are distinguished by the way they estimate the variance of observed processes; the estimate is needed for determining the width of the confidence intervals. Usually, the methods impose special requirements on how the output data from simulation experiments should be collected and preprocessed, which depends on whether they attempt to weaken or even remove statistical dependencies among observations or take the actual correlations among observations into consideration. The following methods can be distinguished:

- The method of independent replications
- The method of batch means
- The method of overlapping batch means

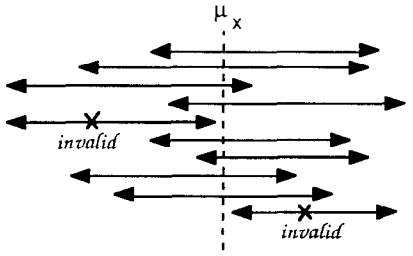

Figure 2. Valid and invalid confidence intervals for μ_x in coverage analysis (in this example the coverage equals 80%).

- The method of uncorrelated sampling
- The method of regenerative cycles
- The method based on spectral analysis
- The method based on autoregressive representation
- The method based on standardized time series

Various approximations assumed during statistical analysis of simulation output data may bias both point and interval estimates. For example, the final confidence interval should theoretically contain the true value of the estimated parameter with probability $(1 - \alpha)$ or, equivalently, if an experiment is repeated many times, in $(1 - \alpha)100\%$ of cases; but various difficulties in satisfying theoretical assumptions can cause the real rate of the confidence intervals containing the true parameter to differ significantly from $(1 - \alpha)$. The robustness of the above methods of data collection and analysis is usually measured by the *coverage of confidence intervals*, defined as the frequency with which the intervals $(\bar{X}(n) - \Delta_x, \bar{X}(n) + \Delta_x)$ contain the true parameter μ_x at a given confidence level $(1 - \alpha)$, $0 < \alpha < 1$; see Figure 2. Thus, the coverage analysis can be applied only to systems with theoretically well-known behavior, since the value of μ_x has to be known. Any analyzed method must be applied in a statistically significant number of repeated simulation experiments (usually 200 or more replications) to determine the fraction of experiments producing the final confidence intervals covering the true mean value of the estimated parameter.

The coverage error and its main sources were theoretically analyzed by Glynn

[1982], Kang and Goldsman [1985], and Schruben [1980]. Generally, a given method of data collection and analysis can be considered as producing valid $100(1-\alpha)\%$ confidence intervals (for, say, the mean delay) if the upper bound of the confidence intervals for the coverage is at least $(1-\alpha)$. Otherwise, confidence intervals for the estimated parameter should be regarded as invalid and the method as inaccurate. A few additional measures for the effectiveness of methods used for data collection and analysis were proposed in Schruben [1981]. The weakest point of such an approach is that there is no theoretical basis for extrapolating results found for simple, analytically tractable systems to more complex systems, which are the real subjects of simulation studies; see Fox [1978].

Some conclusions on the quality of the methods can be also obtained from the analysis of the asymptotic properties of the variance estimators $\hat{\sigma}^2[\bar{X}(n)]$ used by particular methods for determining the width of confidence intervals. Namely, the quality of various variance estimators can be compared by comparing the limit values of their bias,

$$\begin{aligned}&\text{Bias}\{\hat{\sigma}^2[\bar{X}(n)]\}\\&\quad = \mathrm{E}\{\hat{\sigma}^2[\bar{X}(n)] - \sigma^2[\bar{X}(n)]\},\end{aligned} \tag{11}$$

and variance, $\text{Var}\{\hat{\sigma}^2[\bar{X}(n)]\}$, as the number of observations tends to infinity; that is, as $n \to \infty$; see Goldsman and Meketon [1985] and Goldsman et al. [1986]. Alternatively, Schmeiser [1982] proposed studying the asymptotic properties of the expected values and variances of the halfwidth of confidence intervals Δ_x generated by a method; see also Glynn and Iglehart [1985] and Goldsman and Schruben [1984]. Following such criteria, one can say that the method using the variance estimator with the smallest bias and smallest variance, or using the estimator of the width of confidence interval having the smallest expected value and smallest variance, is (asymptotically) superior to others. Unfortunately, these criteria are not universal since a small bias can be accompanied by large variance or vice versa. In terms of confidence intervals, it can mean wide and very variable confidence intervals giving good coverage, or, conversely, stable and narrow confidence intervals giving poor coverage.

Even if we were able to collect independent and identically distributed output data from simulation runs, we cannot be fully protected from erroneous conclusions because of inherent variations of simulation output data caused by the pseudorandom nature of input data; see Pidd [1984, Sec. 8.4.2] for a more detailed discussion. Certainly, one of the most important issues of any simulation experiment is to use proper input data. In the case of queuing processes this usually means selecting a good generator of uniformly distributed (pseudo)random numbers. The state of art in this area is summarized in Park and Miller [1988].

The experimental accuracy of simulation can be further improved by using *variance reduction techniques* (VRT) developed for reducing the variance of recorded results without affecting their mean values. Surveys of variance reduction techniques may be found in Bratley et al. [1983, Sec. 2], Cheng [1986], Frost et al. [1988], Kleijnen [1974, Chap. 3], Law and Kelton [1982a, Chap. 11], Nelson and Schmeiser [1986], Wilson [1983, 1984]. Although all of these techniques can be applied to the method of independent replications, only some of them can be fitted into other methods of data collection and analysis. Two VRTs, known as the method of control variables (or control variates) and importance sampling, seem to be the most frequently advocated; see, for example, Anonuevo and Nelson [1986], Frost et al. [1988], Izydorczyk et al. [1984], Lavenberg and Welch [1981], Lavenberg et al. [1982], and Venkatraman and Wilson [1985], for discussion of these methods and their applications. Unfortunately, despite the fact that various VRTs have been extensively studied theoretically since the beginning of digital simulation [Harling 1958], most of them have found only limited practical application. The reason is that they can be difficult to implement in simulation studies of even moderately complex systems, since they are strongly model dependent and/or

require a substantial amount of computing resources. The simplest VRT in the context of queuing processes [Law 1975], consists of a direct estimation of the mean time in queue (without service time), since this estimator usually has a smaller variance than (direct) estimators of the mean time in the system or the mean queue length. Next, the estimates of the latter group of parameters can be obtained indirectly by applying, for example, Little's formula. The efficiency of this approach has been proved for G/G/c queuing systems [Carson and Law 1980]. A variance reduction as high as 90% has been reported, although generally the reduction tends to 0 as the system utilization tends to 100%. Indirect estimation has been found justifiable even for more complex queuing systems, provided that the interarrival and waiting times are negatively correlated [Glynn and Whitt 1989]. Another VRT, for achieving more stable confidence intervals by reducing the variance $\mathrm{Var}\{\hat{\sigma}^2[\bar{X}(n)]\}$, has been proposed by Meketon and Schmeiser [1984]. They showed that the variance reduction can be achieved by calculating estimators from overlapping subsequences of output data; see the method of overlapping batch means in Section 1.

The methods for data collection and analysis can be used either in their fixed-sample-size versions or in their sequential versions. In the former case, statistical analysis is performed once at the end of the simulation experiment when a predetermined number of observations, assumed to be sufficient to get results of a required accuracy, has been collected. The present survey of methods for data collection and analysis, used in steady-state simulation, focusses on their sequential versions, in which the length of simulation is increased sequentially from one checkpoint to the next until a prespecified accuracy of the point estimators is obtained. Such procedures, which automatically control the length of simulation experiments, are very desirable in user-friendly simulation packages. Sequential statistical analysis is also more efficient, since it is usually difficult to determine a priori the length of simulation needed by a fixed-size procedure that would be sufficient to obtain a required width of confidence intervals at the assumed level of confidence [Law and Kelton 1982b, 1984].

Among the few possible criteria for stopping the simulation, probably the most useful one is based on the relative (half) width of the confidence interval at a given confidence level $(1 - \alpha)$ defined as the ratio

$$\epsilon = \frac{\Delta_x}{\bar{X}(n)} \qquad 0 < \epsilon < 1; \tag{12}$$

c.f. Equation (2).

The above ratio is also called the *relative precision of the confidence interval.* The simulation experiment is stopped at the first checkpoint for which $\epsilon \leq \epsilon_{\max}$, where $\epsilon_{\max}$ is the required limit relative precision of the results at the $100(1 - \alpha)\%$ confidence level, $0 < \epsilon_{\max} < 1$. Note that if

$$1 - \alpha \leq \mathrm{P}[\,|\,\bar{X}(n) - \mu_x\,| \leq \epsilon\,|\,\bar{X}(n)\,|\,], \tag{13}$$

then, for $\mu_x \neq 0$,

$$\begin{aligned}
&\mathrm{P}[\,|\,\bar{X}(n) - \mu_x\,| \leq \epsilon\,|\,\bar{X}(n)\,|\,] \\
&= \mathrm{P}[\,|\,\bar{X}(n) - \mu_x\,| \leq \epsilon\,|\,\bar{X}(n) - \mu_x + \mu_x\,|\,] \\
&\leq \mathrm{P}[\,|\,\bar{X}(n) - \mu_x\,| \leq \epsilon\,|\,\bar{X}(n) - \mu_x\,| + \epsilon\,|\,\mu_x\,|\,],
\end{aligned}$$

and, finally,

$$1 - \alpha \leq \mathrm{P}[\,|\,\bar{X}(n) - \mu_x\,| \leq \epsilon\,|\,\bar{X}(n)\,|\,] \leq \mathrm{P}\left[\frac{|\,\bar{X}(n) - \mu_x\,|}{|\,\mu_x\,|} \leq \frac{\epsilon}{1 - \epsilon}\right], \tag{14}$$

where

$$\frac{|\,\bar{X}(n) - \mu_x\,|}{|\,\mu_x\,|} \tag{15}$$

is called the relative error of the confidence interval.

Sequential reasoning about the statistical accuracy of results is particularly advisable if higher precision is required. There is a danger, however, since when precision requirements are increased, the resulting confidence intervals have a greater chance of not containing the true value of parameters (*the narrower the confidence interval, the worse the coverage effect*). One can also

expect that lower coverage is more probable in the case of negatively correlated observations for which it is more likely to get an underestimated value of their variance. Higher accuracy requirements can also unacceptably lengthen simulation runs controlled by a sequential procedure. In this context, any variance reduction technique can be regarded as a technique for speeding up the simulation, since any decrease in the value of the variance decreases the width of the resulting confidence intervals, and a specified accuracy can be met more quickly.

The sequential approach is regarded by some as the only alternative to steady-state simulation; see Bratley et al. [1983, p. 101]. Some statisticians believe that it is possible to devise a procedure that would automatically conduct data collection and analysis using a sequential rule for assessing the accuracy of estimates; however, a fully acceptable solution has not yet been invented. Relatively few commercially available simulation packages offer some degree of automation of statistical analysis; c.f. Catalog of Simulation Software [1987]. For example, sequential procedures, automated to some extent and based on independent replications, regenerative, and spectral methods of data collection and analysis are implemented in Research Queuing Package (RESQ), [MacNair 1985; Sauer et al. 1984], and in its network-oriented extension Performance Evaluation Tool (PET) [Bharath-Kumar and Kermani 1984]. A method of batch means is incorporated in SIMSCRIPT II.5 [Mills 1987] and its specialized variations such as Network II.5 and COMNET II.5. Partial automation of data analysis is also offered in Queuing Network Analysis Package Version 2 (QNAP2) [Potier 1984, 1986].

The effectiveness of proposed methods generally depends on the level of a priori knowledge of the system's behavior. Successful fully automated implementations have been reported only for some restricted classes of queuing processes. A fully automated procedure that could be used in stochastic simulation studies of a broader class of systems by users having little knowledge or interest in the statistical analysis of the output data is a matter for future investigation. The quality of presently known methods depends on the selection of values for the various parameters involved, which requires some knowledge of the analyzed systems' dynamics. In Section 3 we present details of two sequential procedures for data collection and analysis that were implemented in simulation studies of data communication protocols reported in Asgarkhani and Pawlikowski [1989] and Pawlikowski and Asgarkhani [1988].

This report is not addressed to statisticians. We avoid the strict mathematical formulation of the problems and use basic statistical terminology. Readers are referred to the references for more details.

1. METHODS OF DATA COLLECTION AND ANALYSIS

During the last 25 years of discussion on the methodology of statistical analysis of output data from steady-state simulation, initiated by Conway's [1963] "Some Tactical Problems in Digital Simulation," a variety of methods for data collection and analysis has been proposed to circumvent the nonstationarity of simulated queuing processes (especially the initial nonstationarity caused by the existence of the initial transient period) and the autocorrelation of events (correlations among collected observations). As has been mentioned, these methods either try to weaken (or remove) autocorrelations among observations or to exploit the correlated nature of observations in analysis of variance needed for determining confidence intervals for the estimated parameters.

In **the method of replications**, adopted from terminating simulation, the problem with the autocorrelated nature of the original output data is overcome in a conceptually simple way: The simulation is repeated a number of times, each time using a different, independent sequence of random numbers, and the mean value of observations collected during each run is computed. These means are used in further statistical analysis as secondary, evidently independent and identically distributed output data. By the central limit theorem (see Appendix A), such data also become

approximately normally distributed. Following these motivations, if m observations are collected during each replication, the sequence of n primary observations from $k_b = n/m$ replications $(x_{11}, x_{12}, \ldots, x_{1m})$, $(x_{21}, x_{22}, \ldots, x_{2m}), \ldots, (x_{k_b1}, x_{k_b2}, \ldots, x_{k_bm})$ is replaced by the sequence of their means $\bar{X}_1(m), \bar{X}_2(m), \ldots, \bar{X}_{k_b}(m)$, where

$$\bar{X}_i(m) = \frac{1}{m} \sum_{j=1}^{m} x_{ij}, \tag{16}$$

which are used to obtain the point and interval estimates of the process. Namely, adopting Equations (1)–(4), we get the estimator of the mean μ_x as

$$\bar{\bar{X}}(k_b, m) = \frac{1}{k_b} \sum_{i=1}^{k_b} \bar{X}_i(m), \tag{17}$$

which for $n = k_b m$ is numerically equivalent to $\bar{X}(n)$. We also set the $100(1 - \alpha)\%$ confidence interval of μ_x as

$$\bar{\bar{X}}(k_b, m) \pm t_{k_b-1,1-\alpha/2}\hat{\sigma}[\bar{\bar{X}}(k_b, m)], \tag{18}$$

where

$$\hat{\sigma}^2[\bar{\bar{X}}(k_b, m)] = \sum_{i=1}^{k_b} \frac{\{\bar{X}_i(m) - \bar{\bar{X}}(k_b, m)\}^2}{k_b(k_b - 1)} \tag{19}$$

is the estimator of the variance of $\bar{\bar{X}}(k_b, m)$, and $t_{k_b-1,1-\alpha/2}$, for $0 < \alpha < 1$, is the upper $(1 - \alpha/2)$ critical point from the t-distribution with $k_b - 1$ degrees of freedom.

There are different opinions on the effectiveness of this method as compared to other methods of data collection and analysis, all of which are based on a single (longer) run of the simulation experiment. Arguments defending the method of replications are provided by the results of Kelton and Law [1984], Lavenberg [1981, p. 114], and Turnquist and Sussman [1977], which reveal that better accuracy of the point estimator measured by its MSE [see Equation (7)] can be achieved if the simulation is run a few times rather than if it is run only one time. But Cheng [1976] argues that such a policy cannot always be correct; see also Madansky [1976]. On the other hand, the method of replications appears to be much more sensitive to the nonstationarity of observations collected during the initial transient period than methods based on single simulation runs, since any new replication begins with a new warm-up period. If the bias of the estimator $\bar{\bar{X}}(k_b, m)$ is our main concern, then data collected during the initial transient period should be discarded (see Section 2), and in Equation (17) $\bar{X}_i(m)$ should be replaced by

$$\bar{X}_i(m - n_{oi}) = \frac{1}{m - n_{oi}} \sum_{j=n_{oi}+1}^{m} x_{ij}, \tag{20}$$

where n_{oi} is the number of observations discarded from the ith replication. Thus, the total number of initial observations discarded from k_b replications would be about $k_b - 1$ times larger than in corresponding single run methods. In the sequential version of the method, new replications are generated until the required accuracy is reached. It was found that proper estimation of the length of the initial transient period can significantly improve the final coverage of confidence intervals obtained by the method of replications. There is a trade-off between the number of replications and their length for achieving a required accuracy of estimators. Fishman [1978, p. 122] suggests using at least 100 observations in each replication (i.e., $m - n_{oi} \geq 100$) to secure normality of the replication means. Moreover, results of Law [1977] and Kelton and Law [1984] show that it is better to keep replications longer than to make more replications, since that will usually improve the final coverage too.

All other methods of data collection and analysis have been developed for obtaining steady-state estimators from single simulation runs rather than from multiple replications. In the **method of batch means**, first mentioned by Blackman and Tuckey [1958] and Conway et al. [1959], the recorded sequence of n original observations $x_1, x_2, \ldots, x_n$ is divided into a series of nonoverlapping batches $(x_{11}, x_{12}, \ldots, x_{1m})$, $(x_{21}, x_{22}, \ldots, x_{2m}), \ldots$, of size m, and batch means $\bar{X}_1(m), \bar{X}_2(m), \ldots, \bar{X}_{k_b}(m)$ corresponding to the means over replications from Equation (16) are next used as (secondary) output data in statistical analysis

of the simulation results. The mean μ_x is estimated by $\bar{X}(n) = \bar{\bar{X}}(k_b, m)$ [see Equation (17)], and the confidence interval is given by Equations (18) and (19), and with k_b now meaning the number of batches and m meaning the batch size, $k_b = n/m$. This approach is based on the assumption that observations more separated in time are less correlated. Thus, for sufficiently long batches of observations, batch means should be (almost) uncorrelated; see Brillinger [1973] for a formal justification. By the central limit theorem (see Appendix A), batch means can also be regarded as approximately normally distributed, which justifies the application of Equation (18). If the bias of the estimator $\bar{\bar{X}}(k_b, m)$ is our main concern, then again the effective length of the initial transient period should be determined (see Section 2), and the first n_o observations collected during this period should be deleted. Thus, the division of observations into k_b batches of size m should begin with setting $x_{11} = x_{n_o+1}$.

Selection of a batch size that ensures uncorrelated batch means appears to be the main problem associated with this method. Another problem is selecting a suitable length of the initial transient period. A natural solution is to estimate correlation between batch means starting from an initial batch size m_1, and if the correlation cannot be ignored, increase the batch size and repeat the test. At this stage, the method in its sequential version requires two procedures: the first sequentially testing for an acceptable batch size and the second sequentially testing the accuracy of estimators. Correlation between the means of batches of size m can be measured by estimators of the autocorrelation coefficients

$$\hat{r}(k, m) = \frac{\hat{\mathrm{R}}(k, m)}{\hat{\mathrm{R}}(0, m)} \tag{21}$$

where

$$\hat{\mathrm{R}}(k, m) = \frac{1}{k_b - k} \sum_{i=k+1}^{k_b} [\bar{X}_i(m) - \bar{X}(n)][\bar{X}_{i-k}(m) - \bar{X}(n)] \tag{22}$$

is the estimator of autocovariance of lag, $k = 0, 1, 2, \ldots$, in the sequence of batch means $\bar{X}_i(m), \bar{X}_2(m), \ldots, \bar{X}_{k_b}(m)$. The sequence of batch means can be regarded as nonautocorrelated when all $\hat{r}(k, m)$, $k = 1, 2, \ldots$, assume small magnitudes, say, if they are less than 0.05.

One can also determine the threshold for neglecting the autocorrelations in a statistical way, by testing their values at an assumed level of significance; see Adam [1983] and Welch [1983, p. 306]. The main analytical problem is caused by the fact that $\hat{r}(k, m)$'s of higher order are less reliable since they are calculated from fewer data points.[2] The higher the lag of an autocovariance, the fewer the observations available to estimate this autocovariance within a batch. Usually it is suggested to consider autocovariances of the lag not greater than 25% of the sample size [Box and Jenkins 1970, p. 33] or even of 8–10% (c.f., Geisler [1964]). Law and Carson [1979] have proposed a procedure for selecting the batch size for processes with autocovariances monotonically decreasing with the value of the lag; see also Law and Kelton [1982a]. In such a case, only the lag 1 autocorrelations has to be taken into account. In this procedure three types of behavior of $\hat{r}(1, m)$ as a function of m are distinguished. In the same class of processes Fishman [1978b] has proposed testing batch means against autocorrelation using von Neumann's [1941] statistic. One version of Fishman's procedure can be applied to processes with positive values of $\hat{r}(1, m)$, which decrease monotonically with m, whereas another includes cases when $\hat{r}(1, m)$ is a function oscillating in a damped harmonic fashion, assuming both positive and negative values [Fishman 1978, p. 240]. A sequential procedure using the former version together with the control variates variance reduction technique is presented in Anonuevo and Nelson [1986].

[2] The variance of the estimator $\hat{R}(k, m)$ is reduced if the factor $1/(k_b - k)$ in Equation (22) is replaced by $1/k_b$. But this variance reduction is followed by an increase of the bias of the estimator; see Parzen [1961].

In this procedure observations are batched not by count but by time, that is, over equal time intervals, whose length is specially selected, giving uncorrelated sequence of time means over the intervals.

Procedures proposed for selecting the batch size m^* use various statistical techniques and various criteria, hence they usually lead to very different batch sizes. The statistical tests involved typically require many more batch means to be tested against autocorrelation than is needed for getting results with a required precision. Consequently, it has been reported that some of these procedures can lead to interval estimates with very poor coverage, caused, among other reasons, by accepting batch sizes that are too small. For example, the above-mentioned Fishman's procedures can select batches of as few as eight observations. Law [1983] refers to simulation studies of M/M/1 queues in which the method of batch means with the procedure proposed in Law and Carson [1979] was used. Using $k_b = 10$ batches of size $m = 32$, for system utilization $\rho = 0.9$, and 500 repeated simulation experiments, the achieved coverage of the nominal 90% confidence intervals was only 63%. For these reasons, Kleijnen et al. [1982] suggest the use of a modified Fishman's procedure accepting batches at least 100 observations long, whereas Welch [1983, p. 307] recommends constructing batches at least five times larger than the size m^* given by a test against autocorrelation, provided that at least 10 such batches can be recorded.

Schmeiser [1982] theoretically analyzed the trade-off between the number of batches, the batch size, and the coverage of confidence intervals. The results show that usually the number of batches used in the analysis of confidence intervals should be not less than 10, and need not be greater than 30, if the simulation run is long enough to secure an adequate degree of normality and independency of batch means. This means that having determined a batch size that gives negligibly correlated and approximately normal batch means (which can sometimes require even a few hundred batches to be tested), there is no need to use more than $k_b = 30$ batches to obtain confidence intervals with a good coverage. Thus, confidence intervals can be analyzed by constructing a smaller number of longer batches. Such a transformation improves the normality and independence of batch means and, as such, usually yields better coverage of the confidence intervals. Although the problem of selecting a suitable batch size has not yet been satisfactorily solved, the above method of batch means generally behaves better than the method of replications [Law 1977] and is quite often applied in practice. An example of its sequential version, offering automated analysis of simulation output data, is presented in Section 3.

As mentioned in the introduction, to generate short and stable confidence intervals an estimator of variance $\sigma^2[\bar{X}(n)]$ should have a small variance itself. The theory of statistics says that this requirement is equivalent to using variance estimators with high degrees of freedom, since the number of degrees of freedom is inversely proportional to the variance of such estimators. Meketon and Schmeiser have shown that the variance of the variance estimator can be reduced by introducing overlapping batches of observations [Meketon 1980; Meketon and Schmeiser 1984]. This solution is applied in the **method of overlapping batch means**, a modification of the previous method, which in this context is known as the *method of nonoverlapping batch means.*

Following Meketon and Schmeiser, the variance of $\bar{X}(n)$ is estimated as

$$\hat{\sigma}^2_{\text{ob}}[\bar{X}(n)] = \left(\frac{n}{m} - 1\right)^{-1} \cdot \sum_{j=1}^{n-m+1} \frac{\{\bar{X}_j(m) - \bar{X}(n)\}^2}{n - m + 1}. \tag{23}$$

where

$$\bar{X}_j(m) = \frac{1}{m} \sum_{i=0}^{m-1} x_{j+i} \tag{24}$$

is the batch mean of size m beginning with the observation x_j, and $\bar{X}(n)$ is the overall mean, averaged over all observations, [see

Equation (1)]. Then the confidence interval can be approximated as

$$\bar{X}(n) \pm t_{\kappa,1-\alpha/2}\hat{\sigma}_{\text{ob}}[\bar{X}(n)], \qquad (25)$$

where $t_{\kappa,1-\alpha/2}$ is the upper $(1 - \alpha/2)$ critical point of the t-distribution with $\kappa = 1.5(\lfloor n/m \rfloor - 1)$ degrees of freedom; $\lfloor a \rfloor$ denotes the floor function of a, the greatest integer not greater than a. Thus the number of degrees of freedom is 1.5 times greater here than in the method of nonoverlapping batch means [c.f. Equation (18)], with $k_b = n/m$ nonoverlapping batches. It can be proved that if each new observation starts the next batch of observations, then, assuming the same batch size m, the method of overlapping batch means gives asymptotically (as $n \to \infty$) more stable confidence intervals, since the variance of $\hat{\sigma}^2_{\text{ob}}[\bar{X}(n)]$ equals 2/3 of the variance of $\hat{\sigma}^2[\bar{\bar{X}}(k_b, m)]$ for nonoverlapping batch means [Meketon and Schmeiser 1984]. The bias of both variance estimators remains practically the same [Goldsman et al. 1986]. It has been shown that even overlapping batches only by half of their size gives an asymptotic variance of $\hat{\sigma}^2_{\text{ob}}[\bar{X}(n)]$ equal to 75% of the variance of the estimator calculated from nonoverlapping batches with the same batch size [Welch 1987]. An important feature of this technique is that it makes it possible to increase the size of batches within a given length of simulation run without decreasing the number of batches. All of this makes the estimator given by Equation (23) quite attractive, despite it being computationally more complex, although it remains computationally tractable. The optimal batch size for this estimator (and for others), in terms of its asymptotic mean square error, was analyzed by Goldsman and Meketon [1989] and Song [1988].

Since the method of batch means was originally invented to obtain less correlated data, discarding some observations between consecutive batches should be a natural and effective way to obtain additional decrease in the correlation between batch means. An extreme solution is applied in the **method of uncorrelated sampling** in which only single observations, each of ν observations apart, are retained and all other observations are discarded; see Schmidt and Ho [1988] and Solomon [1983, p. 200]. The distance between consecutive retained observations should be selected large enough to make the correlation between them negligible. When this is done and the initial n_0 observations from the transient period are discarded, the sequence of K retained observations $x_{n_0+1}, x_{n_0+\nu+1}, \ldots, x_{n_0+(K-1)\nu+1}$ contains realizations of (almost) independent and identically distributed random variables. Thus, the mean μ_x can be estimated by

$$\bar{X}_{\text{us}}(K) = \frac{1}{K}\sum_{i=0}^{K-1} x_{n_0+i\nu+1}. \qquad (26)$$

Its confidence interval is

$$\bar{X}_{\text{us}}(K) \pm t_{K-1,1-\alpha/2}\hat{\sigma}_{\text{us}}[\bar{X}_{\text{us}}(K)], \qquad (27)$$

where

$$\hat{\sigma}^2_{\text{us}}[\bar{X}_{\text{us}}(K)] = \sum_{i=0}^{K-1} \frac{\{x_{n_0+i\nu+1} - \bar{X}_{\text{us}}(K)\}^2}{K(K-1)} \qquad (28)$$

and $t_{K-1,1-\alpha/2}$ is the upper $(1 - \alpha/2)$ critical point obtained from the t-distribution with $(K - 1)$ degrees of freedom. The size ν of separating intervals can be selected sequentially, applying the same tests as for determining the batch size in the method of (nonoverlapping) batch means, although one can expect that the size of intervals used for removing correlations between individual observations will usually be smaller than the batch size required for making batch means uncorrelated. In the example considered in Solomon [1983], the separating intervals of length $\nu = 25$ were selected by applying the Spearman rank correlation test. No results on effectiveness of this method are available, but some consider that it wastes too many observations. In fact, as shown by Conway [1963], the benefit of introducing the separating intervals is doubtful since it increases the variance of estimates. Let us also note that one of the reasons for batching observations is to make them more normally distributed. Thus, Equation (27) can give quite a poor approximation to the confidence interval if the analyzed process is not a normal one.

In the **regenerative method** observations are also grouped into batches, but the

batches are of random length, determined by successive instants of time at which the simulated process starts afresh (in the probabilistic sense), that is, at which its future state transitions do not depend on the past. In the theory of regenerative processes (see, e.g., Shedler [1987]), which gives the theoretical support for this method, such instants of time are called *regeneration points*, and the states of the processes at these points are called *regeneration states*. The special nature of the process behavior after each regeneration point—its fresh rebirth—causes batches of observations collected during different *regenerative cycles* (i.e., within periods of time bounded by consecutive regeneration points) to be statistically independent and identically distributed. So are the means of these batches. For example, the regeneration points in the behavior of simple single-server queuing systems are clearly the time instants at which newly arriving customers find the system empty and idle. From any such moment on, no event from the past influences the future evolution of the system. More examples are given in Welch [1983, p. 317] and Shedler [1987, Sec. 2.1]. Note that usually a few, or even infinitely many, different sequences of regeneration points (for different types of regeneration states) can be distinguished in the behavior of a system.

As a consequence of the identical distributions of output data collected within consecutive regenerative cycles, the problem of initialization vanishes if a simulation experiment commences from a selected regeneration point. The regenerative method was first suggested by Cox and Smith [1961, p. 136], then independently developed by Fishman [1973b, 1974], and by Crane and Iglehart [1974, 1975a]. Because of the random length of batches, these methods require special estimators, usually in the form of a ratio of two variables. In particular, if observations $x_1, x_2, \ldots, x_n$ are collected during N consecutive regenerative cycles, then the mean μ_x of the observed process is estimated by

$$\bar{X}_{\mathrm{re}}(N) = \frac{\bar{Y}(N)}{\bar{T}(N)}, \tag{29}$$

where

$$\bar{Y}(N) = \sum_{i=1}^{N} \frac{Y_i}{N}, \tag{30}$$

$$\bar{T}(N) = \sum_{i=1}^{N} \frac{T_i}{N}. \tag{31}$$

In the above equations,

$$T_i = n_{i+1} - n_i \tag{32}$$

is the length of the ith regenerative cycle or, equivalently, the number of observations collected during the cycle i; n_i is the serial number of an observation collected at the ith regeneration point, and

$$Y_i = \sum_{j=n_i}^{n_{i+1}-1} x_j. \tag{33}$$

Thus, Y_i is the sum of observations collected during the ith regenerative cycle. If sufficiently many regenerative cycles are recorded, then the $100(1-\alpha)\%$ confidence interval of unknown parameter μ_x is bounded by

$$\bar{X}_{\mathrm{re}}(N) \pm z_{1-\alpha/2}\hat{\sigma}_{\mathrm{re}}[\bar{X}_{\mathrm{re}}(N)], \tag{34}$$

where $z_{1-\alpha/2}$ is the upper $(1-\alpha/2)$ critical point from the standard normal distribution, and

$$\hat{\sigma}^2_{\mathrm{re}}[\bar{X}_{\mathrm{re}}(N)] = \frac{(s_Y^2 - 2\bar{X}_{\mathrm{re}}(N)s_{YT} + [\bar{X}_{\mathrm{re}}(N)]^2 s_T^2)}{\bar{T}(N)\sqrt{N}}, \tag{35}$$

$$s_Y^2 = \sum_{i=1}^{N} \frac{\{Y_i - \bar{Y}(N)\}^2}{N-1}, \tag{36}$$

$$s_{YT} = \sum_{i=1}^{N} \frac{\{Y_i - \bar{Y}(N)\}\{T_i - \bar{T}(N)\}}{N-1}, \tag{37}$$

$$s_T^2 = \sum_{i=1}^{N} \frac{\{T_i - \bar{T}(N)\}^2}{N-1}. \tag{38}$$

It can be shown that $\bar{X}_{\mathrm{re}}(N)$ given by Equation (29) is a biased estimator of μ_x (the mean value of the ratio of two variables is approximated by the ratio of their mean values, which generally is not correct), although it is a consistent estimator, which means that $\bar{X}_{\mathrm{re}}(N)$ tends to μ_x with probability 1, as $N \to \infty$. Additionally, the asymptotic normality of the ratio estimator

$\bar{X}_{re}(N)$, on which the formula given by Equation (34) is based, is questionable even for relatively large N. Thus, this method eliminates the bias of initialization but introduces new sources of systematic errors caused by special forms of estimators. Some efforts have been made to obtain less biased estimators than those of Equations (29) and (34). Less biased estimators of μ_x have been proposed in Fishman [1977] (Tin's estimator), [Iglehart 1975] (the "jackknife" estimator) and Minh [1987]. Comparative studies reported in Gunther and Wolff [1980], Iglehart [1975, 1978], and Law and Kelton [1982b] show that using the jackknife approach for the mean and variance estimation can significantly improve the accuracy of the estimates, although some question the generality of these results [Bratley et al. 1983, p. 92]. In some reported cases, especially if a small number of regenerative cycles is recorded, the performance of the regenerative method appears to be poor indeed, worse than that of the method of (nonoverlapping) batch means (see Law and Kelton [1982b, 1984]).

An effective sequential, regenerative procedure for output data analysis has been proposed by Fishman [1977]. Because of reservations about the appropriateness of the assumption of the approximate normality of $\bar{X}_{re}(N)$, the procedure is equipped with a statistical test for normality of the collected data (the Shapiro–Wilk test; see Shapiro and Wilk [1965] or Bratley et al. [1983, App. A]). This normality test requires grouping output data (means over observations collected during consecutive regenerative cycles) into fixed size batches. Fishman [1977] proposed using batches containing data collected during at least 100 cycles and increasing the size of batches if the normality test fails. Results presented in Law and Kelton [1982b] show that this method, although rather more complicated numerically because of testing for normality, produces more accurate results in comparison with both a sequential "plain" regenerative method proposed by Lavenberg and Sauer [1977] and a sequential method of (nonoverlapping) batch means proposed by Law and Carson [1979]. A sophisticated modification of the regenerative method was also proposed by Heidelberger and Lewis [1981], who suggest interactive intervention by users in the process of data collection and analysis for achieving better accuracy. The regenerative method of data collection and analysis requires a regeneration state to be well chosen so as to ensure that a sufficient number of output data can be collected for statistical analysis. To satisfy this requirement, a few approximations to the method have been proposed; see Crane and Iglehart [1975b], Crane and Lemoine [1977], Gunther and Wolff [1980], Heidelberger [1979]. Gunther and Wolff [1980] proposed replacing single regeneration states by sets of states and defining (almost) regenerative cycles bounded by entries of the simulated process to such sets of states rather than to a single regeneration state as in the original method. Such modification can lead to even better accuracy of results than that obtained by the original (accurate) regenerative method, at least in the cases reported in Gunther and Wolff [1980]. But users must still select a proper set of (almost) regenerative states, which can sometimes involve substantial preparatory work. This approach certainly deserves to be more thoroughly compared with others. On the other hand, selecting the most frequently occurring regeneration states does not guarantee the best quality of the estimator. For example, Calvin [1988] has shown that such a selection may even result in the estimator with the largest variance. Thus, a general criterion for selecting regeneration states still remains an open question. The random length of regenerative cycles makes the control of the accuracy of results more difficult, since stopping the simulation at a nonregenerative point can cause a substantial additional bias [Meketon and Heidelberger 1982]. Any variant of the regenerative method offers very attractive solution to the main "tactical" problems of stochastic simulation, but it requires a deeper a priori knowledge of the simulated processes.

As has been said, one can also try to take into account the correlated nature of observations when the variance, needed for the analysis of confidence intervals, is

estimated. The simplest, but usually heavily biased, estimator of the variance $\sigma^2[\bar{X}(n)]$ can be obtained directly from Equation (9). Namely,

$$\hat{\sigma}^2[\bar{X}(n)] = \frac{1}{n}\left[\hat{R}(0) + 2\sum_{k=1}^{n-1}\left(1 - \frac{k}{n}\right)\hat{R}(k)\right], \tag{39}$$

where

$$\hat{R}(k) = \frac{1}{n-k} \cdot \sum_{i=k+1}^{n}[x_i - \bar{X}(n)][x_{i-k} - \bar{X}(n)], \tag{40}$$

for $0 \le k \le n-1$. This estimator can be improved by discarding $\hat{R}(k)$'s of higher order since, as has been mentioned discussing the problem of selecting the batch size in the method of batch means, they are less reliable because they are calculated from fewer data points. Since the above formulas assume that the analyzed observations are taken from a stationary process, the estimation should be forwarded by detecting the effective length of the initial transient period to discard initial nonstationary data.

The serial correlation of observations collected during simulation experiments is more effectively exploited in the **method based on spectral analysis**, first proposed by Fishman [1967] and Fishman and Kiviat [1967]. The method assumes that observations represent the stationary and autocorrelated sequence $X_1, X_2, \ldots, X_n$, and shifts their analysis into the frequency domain by applying a Fourier transformation to the autocorrelation function $\{R(k)\}$, $k = 0, 1, 2, \ldots$, yielding the *spectral density function*

$$p_x(f) = R(0) + 2\sum_{j=1}^{\infty} R(j)\cos(2\Pi fj), \tag{41}$$

for $-\infty \le f \le +\infty$ [Brillinger 1981; Jenkins and Watts 1968].

Note that because of the randomness of the collected observations, the spectral density function is a random function too. Comparing Equations (41) and (9) one can see that, for sufficiently large n,

$$\sigma^2[\bar{X}(n)] \cong \frac{p_x(0)}{n}. \tag{42}$$

Thus, the estimator of $\sigma^2[\bar{X}(n)]$ can be obtained from an estimator $p_x(f)$ at $f = 0$. Several techniques have been proposed for obtaining good estimators of the spectral density function $p_x(f)$. Most of them follow classical techniques of spectral estimation based on the concept of *spectral windows* (special weighting functions introduced for lowering the final bias of the estimators) [Fishman 1973a, 1987a; Jenkins and Watts 1968; Marks 1981]. It can be shown that applying a modification of the so-called Bartlett window gives an estimator of the variance from Equation (42) equivalent to that from the overlapping batch means [Damerdji 1978; Welch 1987]. The best results in the sense of coverage were reported by applying the so-called Tukey–Hanning window [Jenkins and Watts 1968; Law and Kelton 1984]. Using this approach, one can determine the confidence interval of μ_x

$$\bar{X}(n) \pm t_{\kappa,1-\alpha/2}\hat{\sigma}_{sp}[\bar{X}(n)], \tag{43}$$

assuming

$$\hat{\sigma}^2_{sp}[\bar{X}(n)] = \frac{\hat{p}_x(0)}{n}, \tag{44}$$

and $t_{\kappa,1-\alpha/2}$ as the upper $(1-\alpha/2)$ critical point of the t-distribution with κ degrees of freedom. The value of κ depends here on the ratio of n/k_{max}, where k_{max} is the value of the upper lag considered in the autocorrelation function $\{R(k)\}$ [Bratley et al. 1983, p. 97; Fishman 1973a]. This approach can sometimes produce quite accurate final results (see Law and Kelton [1984]), but it cannot be regarded as a good candidate for a more user friendly implementation because of its rather sophisticated nature. In particular, there is no definitive method for choosing the parameter κ [Bratley 1983, p. 97; Fishman 1978a, p. 265; Law and Kelton 1984].

The usefulness of spectral windows in reducing the bias of the estimate $\hat{p}_x(0)$ has been questioned in Duket and Pritsker [1978], Heidelberger and Welch [1981a, 1981b], Wahba [1980]. The last three

papers propose estimating $p_x(0)$ from the *periodogram* $\{\Pi(j/n)\}$, $j = 0, 1, \ldots,$ of the sequence of observations $x_1, x_2, \ldots, x_n$. This periodogram is a function of the discrete Fourier transforms $\{A_x(j)\}$ of the observations $x_1, x_2, \ldots, x_n$; namely,

$$\Pi\left(\frac{j}{n}\right) = \frac{|A_x(j)|^2}{n}, \tag{45}$$

and

$$A_x(j) = \sum_{s=1}^{n} x_s \exp\left[-\frac{2\pi i(s-1)j}{n}\right], \tag{46}$$

where[3] $i = \sqrt{-1}$. It can be shown that for $0 < j < n/2$

$$p_x\left(\frac{j}{n}\right) \approx E\left[\Pi\left(\frac{j}{n}\right)\right]. \tag{47}$$

To find an unbiased estimate of $p_x(0)$ the periodogram is transformed into a smoother function, namely, into the logarithm of the averaged periodogram

$$L(f_j) = \log\left\{\left[\Pi\left(\frac{2j-1}{n}\right) + \Pi\left(\frac{2j}{n}\right)\right] \Big/ 2\right\} \tag{48}$$

for $f_j = (4j - 1)/n$. Next, this smoother function (but still not the smoothest one) is approximated by a polynomial to get its value at zero. The whole approach is discussed in detail in Heidelberger and Welch [1981a] together with a method for calculating κ, the number of degrees of freedom needed in Equation (43); see also the Appendix B. Despite a number of approximations involved, the method produces quite accurate results, in particular in terms of coverage.

Heidelberger and Welch [1981a] proposed a sequential version of their method that uses a limited number of (aggregated) output data points instead of a growing number of individual observations, since, as they show, both individual observations and their batch means (of arbitrary size) can be used in the variance analysis. Namely, if n observations are grouped into b batches[4] of m observations each, then for $n = bm$

$$\frac{p_x(0)}{n} = \frac{p_{\bar{X}(m)}(0)}{b}, \tag{49}$$

where, for $-\infty \le f \le +\infty$,

$$p_{\bar{X}(m)}(f) = R(0, m) + 2\sum_{j=1}^{\infty} R(j, m)\cos(2\Pi fj) \tag{50}$$

is the spectral density function of the autocorrelation function $\{R(k, m)\}$ ($k = 0, 1, 2, \ldots$) of the batch means; see Equation (22). This insensitivity of the method to batching the observations allows the batch size to be increased dynamically (starting from $m = 1$), keeping in memory only a limited number of the batch means. A special batching/rebatching procedure is presented in Heidelberger and Welch [1981a, 1981b]. It appears to be an efficient way of limiting the required memory space. A modified version of this method is presented in Section 3. The method has demonstrated a good coverage of confidence intervals in various applications, even if data are collected asynchronously in time [Asgarkhani and Pawlikowski 1989], despite claims based on the basic assumption of discrete Fourier transformation that it should be applied only in simulation experiments in which observations are collected at equally spaced time intervals [Bratley et al. 1983, p. 96].

Another approach for estimating the variance of correlated observations collected during a single simulation run is applied in the **method based on autoregressive representation** developed by Fishman [1971, 1973a, 1978a]. Again it is assumed that after having decided about observations gathered during the initial transient period, the analyzed sequence of observations $x_1, x_2, \ldots, x_n$ represents a stationary process. The main assumption of the method is that such a sequence of originally correlated observations can be

[3] The symbol i has this special meaning only in Equation (46).

[4] The symbol b, instead of k_b, is used here to emphasize the insensitivity of the method to the number of batches.

transformed into a sequence of independent and identically distributed random variables, called its *autoregressive representation.* The autoregressive representation $y_{q+1}, y_{q+2}, \ldots, y_n$ of order q is defined by the transformation

$$y_i = \sum_{k=0}^{q} c_k(x_{i-k} - \mu_x), \tag{51}$$

for $i = q + 1, q + 2, \ldots, n$. It can be shown, [Fishman 1978] that

$$\mathrm{E}[y_i] = 0, \tag{52}$$

$$\mathrm{E}[\bar{Y}(n-q)] = \sum_{i=q+1}^{n} \frac{y_i}{n-q} \approx \mathrm{C}[\bar{X}(n) - \mu_x], \tag{53}$$

and

$$\sigma^2[\bar{Y}(n-q)] \approx \mathrm{C}^2\sigma^2_{\mathrm{ar}}[\bar{X}(n)], \tag{54}$$

where $\mathrm{C} = c_0 + c_1 + \cdots + c_q$, $c_0 = 1$. $\sigma^2[\bar{Y}(n-q)]$, as the variance of the mean of i.i.d. random variables, can be estimated using Equation (4), provided the coefficients $q, c_1, c_2, \ldots, c_q$ are known. The correct autoregressive order q can be determined by examining the convergence of the distribution of a test statistic to an F distribution (also known as the Fisher distribution, or the Snedecor distribution, or the variance-ratio distribution) [Bratley 1983; Hannan 1970, p. 336] or to a χ^2 distribution [Fishman 1978a, p. 251; Hannan 1970, p. 336]. Having selected q, the estimates of the coefficients of $c_1, c_2, \ldots, c_q$ can be found from a set of q linear equations of the form

$$\sum_{i=1}^{q} \hat{c}_i\hat{\mathrm{R}}(k-i) = -\hat{\mathrm{R}}(k), \tag{55}$$

for $k = 1, 2, \ldots, q$, where $\hat{\mathrm{R}}(k)$ is the estimate of the lag k autocovariance from the sequence of original observations $x_1, x_2, \ldots, x_n$ [Fishman 1978a, p. 249]. Next, having determined $\hat{\sigma}^2[\bar{Y}(n-q)]$, one can find the estimate of the variance of $\bar{X}(n)$, since from Equation (54),

$$\hat{\sigma}^2_{\mathrm{ar}}[\bar{X}(n)] = \frac{\hat{\sigma}^2[\bar{Y}(n-q)]}{\mathrm{C}^2}. \tag{56}$$

Finally, assuming that the variable $\sqrt{n}(\bar{X}(n) - \mu_x)/\hat{\sigma}_{\mathrm{ar}}[\bar{X}(n)]$ is governed by the t-distribution with

$$\kappa = n\frac{\mathrm{C}}{2}\left[\sum_{j=0}^{q}(q-2j)\hat{c}_j\right]^{-1} \tag{57}$$

degrees of freedom (see arguments given in Fishman [1978a, p. 252]), the resulting confidence interval of μ_x is determined as

$$\bar{X}(n) \pm t_{\kappa,1-\alpha/2}\hat{\sigma}_{\mathrm{ar}}[\bar{X}(n)]. \tag{58}$$

The main restriction of the last method seems to be the required existence of an autoregressive representation of the simulated process. Results of empirical studies of the method's efficiency published in Fishman [1971] were not very encouraging, since the final coverage was frequently below 80%. These results were, however, achieved in short simulation runs. Andrews and Schriber, in their studies of the autoregressive method reported in Andrews and Schriber [1978] and Schriber and Andrews [1979, 1981], observed a significant variability in the average widths of confidence intervals in simulation experiments. Law and Kelton [1984], after comparative studies of different fixed-size methods of data analysis, also found that the autoregressive approach does not offer better results than other, computationally simpler methods of data analysis. And, in contrast to both the method of batch means and the method of spectral analysis, the improvement of the final coverage when increasing the number of collected observations was very slow. Continuous execution of the test for determining the autoregressive order q and solving the sets of equations for determining the coefficients $c_1, c_2, \ldots, c_q$ can be time consuming in a sequential version, especially if longer sequences of observations have to be collected.

The **method of standardized time series**, originally proposed by Schruben [1983, 1985], relies on the convergence of standardized random processes to a Wiener random process with independent increments, also known as a Brownian bridge process. It is an application of the theory of dependent random processes [Billingsley 1968, Chaps. 20 and 21] and its functional central limit theorem, which is a generalization of the (scalar) central limit theorem

presented in Appendix A. According to this approach, an analyzed sequence of observations is first divided into subsequences (batches) of observations, and each of them is transformed into its standard form required by the functional central limit theorem. Next, various functions of the transformed sequence can be analyzed to construct the confidence interval of $\bar{X}(n)$. The method requires that the analyzed process be stationary; thus, initial observations representing its nonstationary warm-up period should be discarded before the sequence of n remaining observations is divided into k_b nonoverlapping batches, each of size m. The ith batch, containing the observations $x_{(i-1)m+1}, x_{(i-1)m+2}, \ldots, x_{(i-1)m+m}$, $i = 1, 2, \ldots, k_b$, is transformed into the standardized process $T_i(t)$, $0 \le t \le 1$, where

$$T_i(t) = \frac{\lfloor mt \rfloor [\bar{X}_i(m) - \bar{X}_i(\lfloor mt \rfloor)]}{\hat{\sigma}[\bar{X}_i(m)]\sqrt{m}} \tag{59}$$

for $0 < t \le 1$ ($\lfloor a \rfloor$ denotes the greatest integer not greater than a), and $T_i(0) = 0$. In this formula,

$$\bar{X}_i(k) = \frac{1}{k}\sum_{j=1}^{k} x_{(i-1)m+j} \tag{60}$$

is the cumulative average of the first k observations in the ith batch, and $\hat{\sigma}^2[\bar{X}(m)]$ is the variance estimator of the ith batch mean. The functional central limit theorem says that in the limit, as $m \to \infty$, any standardized process $T_i(t)$, $0 \le t \le 1$, for $i = 1, 2, \ldots$, becomes the Brownian bridge process (a mathematical model of Brownian motion on the [0, 1] interval [Billingsley 1968]). This fact has been used by Schruben [1983], who proposed to use two functions of $T_i(t)$ to estimate the variance of $\bar{X}(n)$: (i) the maximum of $T_i(t)$, $0 \le t \le 1$, and (ii) the sum of $T_i(k/m)$, from $k = 1$ to m. The former function is used in the *maximum estimator* of $\sigma^2[\bar{X}(n)]$,

$$\hat{\sigma}^2_{\max}[\bar{X}(n)] = \sum_{j=1}^{k_b} \frac{k_{i,\max}}{3k_b^2(m - k_{i,\max})} [\bar{X}_i(m) - \bar{X}_i(k_{i,\max})]^2, \tag{61}$$

where

$$k_{i,\max} = \min\{k : k[\bar{X}_i(m) - \bar{X}_i(k)] \ge j[\bar{X}_i(m) - \bar{X}_i(j)]; \quad j = 1, 2, \ldots, m\} \tag{62}$$

is the location of the (first if ties occur) maximum of $T_i(k/m)$ as a function of k, Goldsman and Schruben [1984], and gives the confidence interval of μ_x,

$$\bar{X}(n) \pm t_{3k_b,1-\alpha/2}\hat{\sigma}_{\max}[\bar{X}(n)], \tag{63}$$

where the constant $t_{3k_b,1-\alpha/2}$ is from the t-distribution with $3k_b$ degrees of freedom. The latter function is used in the *area estimator*

$$\hat{\sigma}^2_{\mathrm{A}}[\bar{X}(n)] = \frac{12}{(m^2-1)n^2}\sum_{i=1}^{k_b} \mathrm{A}_i^2, \tag{64}$$

where

$$\mathrm{A}_i = \sum_{j=1}^{m} (j - 0.5(m+1))x_{(i-1)m+j} \tag{65}$$

[Song 1988]. It gives the confidence interval

$$\bar{X}(n) \pm t_{k_b,1-\alpha/2}\hat{\sigma}_{\mathrm{A}}[\bar{X}(n)]. \tag{66}$$

Goldsman and Schruben [1984] showed that the former estimator is asymptotically superior to the later one in the sense that as $m \to \infty$, it produces narrower and more stable confidence intervals. On the other hand, it can perform poorly when batches are short. They also showed that if m is large, the method of standardized time series is superior to the method of nonoverlapping batch means. Another positive feature of this method is that despite the sophisticated statistical techniques involved, the estimators have simple numerical forms. For a further improvement of their quality, Damerdji [1987] proposed to use overlapping batches of observations as is done in the method of overlapping batch means, and Foley and Goldsman [1988] have modified Equation (59), introducing the so-called orthonormally weighted standardized time series, which gives asymptotically narrower and more stable confidence intervals than the original one. Unfortunately, no simple rule for selecting

the batch size is available, and only a few practical implementations of the method of standardized time series have been reported. Theoretical analysis of asymptotic cases shows that the method usually requires longer batches than the method of nonoverlapping or overlapping batch means [Song 1988].

The estimators of variance that have been presented in this section require various, more or less complex ways of data collection. They also have their own statistical strengths and weaknesses. This has led to the idea of finding a robust method of simulation output data analysis by combining different estimators of variance $\sigma^2[\bar{X}(n)]$ into a composite estimator, since, by theoretical arguments, combinations of independent variance estimators should have a smaller variance and, consequently, give better coverage of confidence intervals than its components. Such an effect has been achieved by Schruben [1983], with a linear combination of the area estimator (64), or the maximum estimator (61), and the estimator calculated from nonoverlapping batch means; see results of the quality analysis in [Goldsman and Schruben 1984; Goldsman et al. 1986; Song 1988]. The first of these two combined estimators has been applied in a sequential procedure for the analysis of simulation output data described in Duersch and Schruben [1986]. More sophisticated linear combinations of estimators are discussed in Schmeiser and Song [1987] and Song and Schmeiser [1988].

Research in the area of variance estimation continues. The diversity of existing methods of data collection and analysis, and the variance estimators they use, requires a more thorough comparison of their quality. Some results of comparative studies can be found in Glynn and Iglehart [1985], Goldsman and Meketon [1985, 1989], Goldsman and Schruben [1984], Goldsman et al. [1986], Kang and Goldsman [1985], and Song [1988]. Readers interested in current developments in this area are encouraged to browse through recent publications in, for example, "Operations Research," "Management Science," and the annual *Proceedings of the Winter Simulation Conference.*

2. PROBLEM OF INITIALIZATION

It is well known that just after initialization any queuing process with nondeterministic, random streams of arrival and/or random service times is in a transient phase, during which its (stochastic) characteristics vary with time. This is caused by the fact that, like any (stochastic) dynamic system, such queuing systems or networks initially "move" along nonstationary trajectories. After a period time, the system approaches its statistical equilibrium on a stationary trajectory if the system is stable, or remains permanently on a nonstationary trajectory if the system is unstable.[5] Note that in practice only queuing systems with infinite populations of customers and unlimited queue capacities can never enter a stationary trajectory, and this happens if the average request for service is equal to or greater than the average supply of service; that is, if

$$\lambda \geq c\mu_s, \tag{67}$$

where λ is the mean arrival rate, $1/\mu_s$ is the mean service time, and c is the number of service facilities. In such a case, the queues will eventually increase in length with time and the system becomes permanently congested. On the other hand, queuing systems with limited queue capacities always reach an (inner) statistical equilibrium, even if the system's load expressed by the traffic intensity $\rho = \lambda/c\mu_s$ is much greater than 1. In such a case, internally stationary queuing systems are in the nonstationary environment of streams of rejected customers. Of course, output data collected during transient periods do not characterize steady-state behavior of simulated systems, so they can cause quite significant deviation of the final "steady-state" results from their true values. Although it seems quite natural that the deletion of untypical initial observations should result in better

[5] We could also mention the periodic limiting behavior of D/D/c queuing systems and their networks.

steady-state estimators, the problem to delete or not to delete is a perennial dilemma of stochastic simulation practice. Each of these two alternatives has its advocates. The answer depends on the assumed measure of goodness and the resource limitations of simulation experiments (the maximum possible number of recorded observations). The influence that the initial transient data can have on the final results is a function of the strength of the autocorrelation of collected observations. With no restrictions imposed on the length of the simulation run, this influence can be arbitrarily weakened by running the simulating program sufficiently longer. But in most practical situations simulation experiments are more or less restricted in time, and that time can be more or less effectively used to calculate estimators. If all initial output data are retained, the bias of the point estimator $\bar{X}(n)$ is greater than if they were deleted.

Contrary opinions on the usefulness of deletion are caused by the fact that it increases the variance of the point estimator [Fishman 1972, 1973a, Sec. 10.3; Turnquist and Sussman 1977] and, in effect, can increase its MSE; see Equation (8). Let us note that an increase of the variance can be compensated for by applying one of the variance reduction techniques. Deletion of initial observations seems to be justified if the variance of the estimator is smaller than the squared bias and/or if observations are strongly correlated (the initial conditions have a longer effect on the evolution of the system in time). On the other hand, Blomqvist [1970] showed that for long run simulations of GI/G/1 queuing systems the minimum MSE of the mean delay usually occurs for the truncation point $n_0 = 0$, which supports the thesis that no initial observations need to be deleted. Results of experiments conducted by Turnquist and Sussman [1977] and Wilson and Pritsker [1978b] provide the same argument.

The usefulness/uselessness of data deletion also depends on methods used for data collection and analysis. Independent replications give much more "contaminated" data than methods of data collection based on single runs, since each replication begins with a new initial transient period. Consequently, data deletion seems to be more crucial for plurality of transient periods than for just one transient period in one long run. In an example discussed in Kelton and Law [1983] the estimator of mean delay in an M/M/1 queue obtained from replications of 500 observations, without initial deletions, was biased −43.2%, for $\rho = 0.95$. In the case of higher accuracy requirements, a significant bias of estimators will normally increase their chances of being outside the theoretical confidence intervals, thus it will decrease the coverage of confidence intervals. Law [1983] and Kelton and Law [1984] analyzed the influence of initial data deletion on the coverage of the final results in the case of the method of independent replications and stated a clear improvement of the actual coverage (experimental values of confidence levels) to levels near nominal theoretical values $(1 - \alpha)$, without unduly widening confidence intervals, especially if replications were not too long and/or not too many observations were deleted. In methods of data analysis based on single runs and an assumption that the observed process is stationary, deletion of data from the initial nonstationary period improves approximate stationarity of the remaining process.

The nature of the convergence of simulated processes to steady state depends on many factors; the initial conditions of simulation are one of them. Conway [1963] advised a careful selection of starting states (typical ones for steady state of the simulated process) to shorten the duration of the initial transient phase. Since then many trials have been undertaken to determine the optimal initial conditions in the sense that they would cause the weakest influence of the transient phase on the steady-state results, but ambiguous conclusions have been reached. Madansky [1976] proved that the MSE of the mean queue length in simulation studies of M/M/1 queuing systems (without data deletion) can reach its minimum value if they are initialized as empty and idle, that is, in their modal states. Wilson and Pritsker [1978b], having examined a slightly broader

class of queuing processes, concluded that the optimal (in the MSE sense) initial state is the most likely state in statistical equilibrium (the mode of the steady-state distribution) if it differs from the empty-and-idle state. Moreover they found that a judicious selection of initial conditions can be more effective than the deletion of initial data. Similar conclusions were reached by Donnelly and Shannon [1981] after a more methodical investigation. Following this line of reasoning, Murray and Kelton [1988] proposed to use the method of independent replications with randomly selected initial states taken from an initial state distribution. Such a distribution could be constructed during the first k_0 replications (the pilot phase of the simulation experiment) and used during the remaining $k_b - k_0$ replications (the productive phase). Results reported in Murray and Kelton [1988] show that such an approach can be effective both for reducing the bias of point estimates and for increasing the coverage of confidence intervals, without unduly increasing the variance or mean square error of estimates.

On the other hand, Kelton and Law showed experimentally by investigating queuing systems with exponential and Erlangian distributions of interarrival and service times that the shortest transient periods occur if the simulated processes start from states slightly larger than their steady-state means [Kelton 1985; Kelton and Law 1985]. For example, in the M/M/1 queue the mean delay reaches its steady state in the shortest time (with accuracy ± 0.01, $\rho = 0.9$) if the initial queue length is 15, while the steady-state mean queue length equals 9. These results were theoretically justified by Abate and Whitt [1987b], who indicated that the optimal initial state when the mean value is estimated is about one-and-a-half times the steady-state mean. It was also shown that starting from a state much larger than the mean can result in a very long transient period. Thus, because in real situations the steady-state mean is unknown, it is much safer to initialize systems as empty and idle, particularly if the bias of an estimator concerns us more than its MSE.

Having decided to discard data collected during transient periods, we face the next problem: how long such periods last. In simulation practice we can encounter both very short initial transient effects and transient effects that are spread over tens of thousands of observations [Heidelberger and Welch 1983].

2.1 Duration of the Initial Transient Period

The problem of determining the duration of the initial transient period in simulation runs appears to be complicated, even if we restrict ourselves to estimators of means values only. The first rules of thumb were proposed by Conway [1963] and Tocher [1963]. Conway suggested the following rule:

R1. In a time series of observations x_1, x_2, ..., x_n the initial transient lasts until the first of the series is neither the maximum nor minimum of the rest.

This rule of thumb, associating the beginning of steady state with the occurrence of the first "typical" observation, appears to give a poor approximation of the duration of the initial transient. As was shown in Gafarian et al. [1978], using this rule we can significantly overestimate the length of the initial transient for small ρ and underestimate it for high ρ; see also [Wilson and Pritsker 1978b].

The performance of a system can be regarded as a cyclic evolution of system's basic operations. For this reason, Tocher [1963, p. 176] suggested this rule:

R2. The initial transient period is over if the longest cycle distinguished in the behavior of the simulated system has been executed at least three or four times.

No results concerning the effectiveness of this rule are available. The duration of the initial transient period is also analyzed in the queuing theory. It has been shown that the rate at which the mean queue lengths or the mean delays tend to their steady state is, after some period of time, dominated by a term of the form $\exp(-t/\tau)$, where τ is called the *relaxation time* of the queue. Thus, the constant τ

may be used for specifying an upper bound on the length of time after which the influence of the initial state is negligible. For example, one can conclude this rule:

R3. The initial transient period is over after the time $t_\beta = -\tau \ln \beta$, *where* β *is the permissible relative residue of the initial state,* $0 < \beta < 1$.

Thus, assuming $\beta \leq 0.02$, at $t = 4\tau$ we find that the queue characteristics are within 2% of their steady-state values; or in other words, output data collected from that point of time should be biased by initial states by less than 2%. The analysis of relaxation times was initiated by Morse [1955], who considered the correlation function of the M/M/1 queue length. Cohen [1982] analyzed transient distributions of queue lengths and determined the relaxation time for GI/G/1 queuing systems. These appear to be from 9 to 2 times greater than Morse's results for M/M/1 systems, as ρ changes from 0.1 to 1.0. This diversity of results has stimulated search for approximate formulas for relaxation times, such as Newell's result for GI/G/1 under heavy traffic queues [Newell 1971] and results for Markovian queuing systems obtained by Odoni and Roth [1983]. The latter, having studied various Markovian systems, proposed to approximate the relaxation times by

$$\tau = \frac{C_a^2 + C_s^2}{2.8\mu_s(1 - \sqrt{\rho})^2}, \tag{68}$$

where C_a^2 and C_s^2 are the coefficients of variation for the interarrival and service times, respectively, and $1/\mu_s$ is the mean service time. The usefulness of the last formula in simulation has been studied in the case of M/M/1 and $M/E_k/1$ queuing systems [Roth 1985; Roth and Rutan 1985]. More detailed analysis of the transient behavior of some stochastic processes has been reported by Abate and Whitt [1987a, 1987b, 1987c, 1988], who analyzed the relaxation times in the M/M/1 queues and Brownian motion processes (the latter are used to approximate queuing processes in a heavy traffic scenario [Kleinrock 1976]). All these results show that more heavily loaded systems tend more slowly to their statistical equilibrium. Abate and Whitt have also proved that higher moments of queue parameters have longer relaxations times than corresponding lower moment [Abate and Whitt 1987a]. Thus, mean values tend to a steady state faster than, for example, variances.

Relaxation times have also been analyzed theoretically for some simple queuing networks; for example, Blanc [1985a] analyzed the relaxation time in an open network of K service centers with a Poisson arrival stream, an unlimited number of servers at each center, general distribution of service times, and a homogenous transition matrix. He showed that the relaxation time in such a network has an upper limit, namely,

$$\tau \leq \frac{K}{\mu_s}. \tag{69}$$

Equality occurs for tandem connections of queuing systems. The case of two queuing systems in series has been analyzed in more detail in Blanc [1985b].

In addition, a conjectural relaxation time for Jacksonian queuing networks with K single server centers has been proposed. For more complex queuing networks the relaxation times have not yet been theoretically determined. But the usefulness of even known formulas for relaxation times can be questioned in simulation studies. They can be used only as first approximations of the duration of simulated initial transients, since it has been shown that estimators of the mean values from simulation tend to their steady state more slowly than exponentially; for example, Anderson [1985] showed that in queuing systems with limited queue capacities the rate at which the estimator of mean queue length tends to its steady state eventually becomes inversely proportional to time. It has also been shown that the standard deviation of estimators converges even more slowly, namely, in inverse proportion to the square root of time [Anderson 1985; Fishman 1967]. Both of these facts have found application in various heuristic rules proposed for determining the duration of the initial transient period.

Studying the convergence of a moving average of output data to determine a possible end of the initial transient period is attributable to Gordon [1969, p. 285] and Emshoff and Sisson [1970, Sec. 8.2]; see Solomon [1983, p. 195]. The simplest way would be to find an instant of time at which the running mean $\bar{X}(n)$ [Equation (1)] approaches a constant level with a given accuracy δ, $\delta > 0$. Thus, we can assume the following rules:

R4. In a time series of observations $x_1, x_2, \ldots, x_i, \ldots$, *the initial transient period is over after* n_o *observations if* k *consecutive values of the running mean* $\bar{X}(i)$ *recorded after the observation* n_o *differ less than* $100\delta\%$ *from* $\bar{X}(n_o + k)$; *that is, for all* i, $n_o < i \leq n_o + k$,

$$\frac{|\bar{X}(n_o + k) - \bar{X}(i)|}{|\bar{X}(n_o + k)|} < \delta. \qquad (70)$$

The stabilization of $\bar{X}(n)$ should be tested over a sufficiently long sequence of observations, so the parameter k should be large (in the statistical sense); that is, $k \geq 30$. The above rule has two weaknesses. First, as has been indicated by Conway [1963], accumulative statistics such as running means usually stabilize very slowly with time, and therefore usually give overestimated values of n_o. Second, fluctuations of the running mean $\bar{X}(n)$, calculated over data collected during a single simulation run, can continue for a long time. For these reasons the above rule and its various modifications are usually used with the method of replications, and the inequality (70) is applied to the running mean after having additionally smoothed it by averaging over replications. Despite this, the resulting length of the initial transient period is still usually overestimated [Gafarian et al. 1978; Roth 1985; Roth and Rutan 1985; Wilson and Pritsker 1978a]. Welch proposed a special technique for smoothing running mean that uses the concept of a moving window within which mean values over replications are additionally averaged, producing a smoother (but still highly correlated) process [Welch 1983, p. 293]. No results on the effectiveness of this technique have been published.

Another rule of thumb can be based on the supposition that in steady state typical observations are evenly dispersed around the mean value. For example, Fishman [1973b, p. 275] proposed the following:

R5. The initial transient period is over after n_o *observations if the time series* $x_1, x_2, \ldots, x_{n_o}$ *crosses the mean* $\bar{X}(n_o)$ k *times.*

This rule is sensitive to the value of k [Gafarian et al. 1978]. Too large a value will usually lead to an overestimated value of n_o regardless of system's utilization, whereas too small a value can result in an underestimated n_o in more heavily loaded systems. In Gafarian et al. [1978], $k = 25$ was recommended for M/M/1/∞ queuing systems, whereas in Wilson and Pritsker [1978b] $k = 7$ was chosen for the M/M/1/15 system. The system-dependent selection of the parameter k in rule R5 seems to be too arduous for potential users.

Yet another approach, which Solomon [1983] attributes to Emshoff and Sisson [1970], is based on the χ^2 goodness-of-fit test applied for selecting a time from which the numbers of observations below and above the running mean are equal (in the statistical sense). According to this test, the sequence of observations should be partitioned into batches of at least $m_o = 10$ observations each (Solomon selected $m_o = 30$). Then one can conclude the following rule:

R6. In a time series of observations $x_1, x_2, \ldots, x_n$ *the initial transient is over after* n_o *observations if the* χ^2 *goodness-of-fit test confirms that in the batch of observations* $x_{n_o+1}, x_{n_o+2}, \ldots, x_{n_o+m_o}$ *following the observation* n_o *the numbers of observations above and below the running mean* $\bar{X}(n_o)$ *are about the same.*

Rule R6 seems to be quite simple and independent of any system-related parameter. No results are available on its effectiveness and relationship to other criteria. For reducing the fluctuations of analyzed sequences and saving memory space if long transient periods are expected, batches of individual observations can be replaced by their mean values. Note that such batches

must be introduced before the procedures for selecting the size of uncorrelated batch means discussed in Section 1 can be applied. There are no established rules in this case for selecting the batch size. If a statistical test is used to help decide about the length of the initial transient, we should follow the requirements of the test or use statistically large batches, which usually means taking $m_o > 30$. Otherwise, the only recommendation is to select a batch size that gives the desired data reduction while retaining the stabilizing trend of the original sequence. After the batch size m_o is selected, the sequence of batch means $\bar{X}_1(m_o)$, $\bar{X}_2(m_o)$, ..., can be tested in a similar way to the sequence of the original observations. For example, Wilson and Pritsker [1978a] formulated the following rule, which they attributed to Schriber [1974] (Solomon [1983, p. 195] attributed it to Emshoff and Sisson [1970]):

R7. In a time series of batch means $\bar{X}_1(m_o)$, $\bar{X}_2(m_o)$, ..., *the initial transient is over after* b_o *batches, that is, after* $n_o = b_o m_o$ *observations, if the k most recent batch means all fall within an interval of width* δ_1; *that is, if*

$$| \bar{X}_{b_o-i}(m_o) - \bar{X}_{b_o-j}(m_o) | < \delta_1, \quad (71)$$

for $0 < i < k - 1$, $0 < j < k - 1$.

This rule, like rule R4, is sensitive to the value of the parameter k, which should depend on the variability of the observed process. A small value for k, for example $k = 2$, as was assumed in [Wilson and Pritsker [1978a] and Solomon [1983, p. 196], can lead to an underestimation of n_o, since the difference between averages, having dropped only k times below δ_1, can easily rise again to an unacceptable level, as in an example considered in Solomon [1983, p. 197].

For further data reduction and additional smoothing of the tested sequence, Kelton and Law [1983] proposed applying such a batching technique in connection with the method of independent replications; see Section 1. Namely, they batched the sequence of mean observations (means over replications) and then analyzed the sequence of the means of these batches, assuming rule R8:

R8. In a time series $\bar{X}_1(m_o)$, $\bar{X}_2(m_o)$, ..., *the initial transient period is over after the batch* b_o, *that is, after* $n_o = m_o b_o$ *observations, if the sequence of the batch means after the batch* b_o *can be approximated by a straight line with zero slope.*

Rule R8 can be applied only in the case of monotonic convergence to steady state but, as was proved by Kiefer and Wolfowitz [1955], in any stable, initially empty-and-idle GI/G/c queuing system the mean delay in queue grows monotonically in time. Kelton and Law [1983] proposed testing the slope of the regression line backwards after collecting an assumed number of observations. The test for zero slope is over a fixed number of batch means (if zero slope is confirmed, the test is repeated over an earlier sequence of batch means to find whether the initial transient period had expired earlier). If the test fails at the beginning, a new checkpoint is chosen after gathering further output data from the simulation. Note that this requires that the process of collecting new observations in all previously stopped replications be continued again. Because correlations between batch means can still be significant, they are approximated by a straight line using a generalized least-squares procedure proposed by Amemiya [1973], which allows for autocorrelation of the analyzed data. For additional saving of memory space, the number of batches could be kept constant by allowing the size of batches to grow when simulation runs are continued. The procedure implementing rule R8 appears to be quite effective, especially in lowering the MSE of estimators [Roth 1985; Roth and Rutan 1985].

Rules R4–R8 are based on the convergence of the mean of observations to its steady-state value. Other criteria of convergence are also possible. For example, because the variance of the mean of observations taken from a stationary process is approximately inversely proportional to the number of observations [Fishman 1973a, p. 281; Gafarian et al.

1978]; that is,

$$\hat{\sigma}^2[\bar{X}(n)] = \frac{C_1}{n} + o\left(\frac{1}{n}\right) \qquad (72)$$

[cf., Equation (9)] where C_1 is a positive constant and n is the number of observations, Gordon [1969] proposed rule R9:

R9. In a time series $x_1, x_2, \ldots, x_n$ the initial transient is over after the observation n_o if the graph (log n, log $\hat{\sigma}[\bar{X}(n)]$), *becomes approximately linear with slope* -0.5 *from this observation on.*

To smooth variations of the analyzed curve, Gordon [1969] proposed analyzing the variance of the mean of observations averaged over a number of replications. This rule was analyzed in Gafarian et al. [1978] and Wilson and Pritsker [1978a], using Equation (4) to calculate $\hat{\sigma}^2[\bar{X}(n)]$, thus rejecting existing correlations between observations. In this case rule R9 can give an overestimated value of n_o. No results have been published on the effectiveness of this rule when more accurate estimators of $\sigma^2[\bar{X}(n)]$ are applied.

Fishman [1971, p. 29] proposed equating the variance of the mean of autocorrelated observations with the variance of the mean of a hypothetical sequence of independent observations to find the number of collected (autocorrelated) observations equivalent, in the above sense, to one independent (hypothetical) observation. After some simplification we get the following rule:

R10. In a time series of observations $x_1, x_2, \ldots, x_n, \ldots$, the initial transient is over after

$$n_o = 2 \sum_{k=1}^{n-1} \left(1 - \frac{k}{n}\right) \frac{\hat{R}(k)}{\hat{R}(0)} \qquad (73)$$

observations, where $\hat{R}(k)$ is the estimator of the autocorrelation of the lag k, $0 \le k \le n - 1$; see Equation (40).

The sequence of observations collected after the observation n_o should be (approximately) independent of the initial conditions. The autocovariance estimators $\hat{R}(k)$ should be analyzed with caution; see the comments after Equation (22). Comparing this with the results given for the example in Roth and Rutan [1985], one can state that Rule R10 usually gives underestimated values of n_o; no exhaustive comparisons of this rule with other rules are available.

As an example of more sophisticated truncation rules let us mention the rule proposed in Beall [1982] based on the analysis of the convergence of the distribution functions to their steady-state forms. The rule has been analyzed in the context of autoregressive moving average (ARMA) processes [Brockwell and Davis 1987, Chap. 3]. All the above-mentioned rules proposed for determining the length of the initial transient periods are either quite elaborate and, as such, do not ensure an accurate control of the initialization bias, or can determine precisely the length of the initial transient period but only for restricted classes of simulated processes and/or by applying sophisticated techniques to collect and analyze the output data. This can unnecessarily lengthen the time of simulation experiments, especially if the required accuracy is tested sequentially; cf. rule R8. Usually the effectiveness of these rules also strongly depends on the specific parameters they use, but little or no guidance is available on how to select values for these parameters. Some of the rules have been implemented as built-in options offered in simulation packages such as GPSS, SLAM, and SIMSCRIPT II.5 [Law and Kelton 1982a]. Thus, potential users should be aware of their limitations.

A promising approach for detecting the expiration of the initial transient period is offered by statistical stationarity tests based on the theory of dependent stochastic processes developed by Billingsley [1968]. According to this approach, we have the following rule: R11:

R11. The initial transient data have been removed from a given sequence of observations if the (standardized) sequence determined over the remaining observations behaves in a way consistent with a standard (stationary) stochastic process.

Such tests were first invented by Schruben [1982] and Schruben et al. [1983] as a special application of the same theory that is applied in the method of standardized time series for generating confidence intervals; see Section 1. These tests are based on the high sensitivity of the sequence of partial sums

$$S_k = \bar{X}(n) - \bar{X}(k), \qquad (74)$$

$k = 0, 1, 2, \ldots$, and $S_0 = S_n = 0$, to the presence of initialization bias in $\bar{X}(n)$; $\bar{X}(n)$ and $\bar{X}(k)$ are means over n and k first observations, respectively; see Equation (1). Following this phenomenon, tests proposed in Schruben [1982] and Schruben et al. [1983] analyze the convergence of a standardized sequence $\{T(t)\}$, $0 \le t \le 1$, to the Brownian bridge process with zero mean and variance equal 1. The sequence $\{T(t)\}$ is the standardized sequence of the partial sums S_k, namely,

$$T(t) = \frac{\lfloor nt \rfloor S_{\lfloor nt \rfloor}}{\hat{\sigma}[\bar{X}(n)]\sqrt{n}}, \qquad (75)$$

for $0 < t \le 1$ (where $\lfloor a \rfloor$ denotes the greatest integer not greater than a), and $T(0) = 0$.

Heidelberger and Welch [1983] listed a few other standardized sequences that can be used to find statistics for the above rule. Rejection or acceptance of the hypothesis that a given subsequence of observations is stationary or, equivalently, that the initial transient period is not included in the observations, depends on the probability characterizing the scalar value calculated from the considered sequence. Despite the sophisticated theory hidden behind these tests, they appear to be simple numerically and can be applied to a wide class of simulated processes. A sequential version of one of the tests proposed by Schruben et al. [1983] is presented in Section 3. The main practical problem with their implementations is that they require a priori knowledge of the variance estimator $\hat{\sigma}^2[\bar{X}(n)]$ of the simulated process in steady state. To estimate this variance, one can use a sequence of observations collected at some distance from an assumed truncation point, assuming that the process is then at least closer to steady state. Schruben [1982] also describes a test that does not require the variance $\sigma^2[\bar{X}(n)]$, suggesting that it should perform not worse than others. No results of its evaluation are available. Note that searching for the beginning of the stationary phase by means of any statistical test can be shortened if it is preceded by one of the rules of thumb, for example, rule R1 or R6, to find a rough approximation of the length of the initial transient period.

Recently, Vassilacopoulos [1989] proposed a new, simple statistical test, which does not require the calculation of the variance of the simulated process. But as its author stated, the test should be evaluated extensively before it is adopted into simulation practice.

3. SEQUENTIAL PROCEDURES FOR STEADY-STATE SIMULATION: EXAMPLES

This section presents in detail two sequential procedures for stopping the simulation experiment when the required relative precision of confidence intervals is achieved. The first procedure is based on the method of spectral analysis; the second applies the method of the nonoverlapping batch means. Since both of these procedures require the analyzed sequence of observations to be stationary, each of them has to be preceded by a sequential procedure for detecting the length of the initial transient (nonstationary) period. Thus, when these procedures are applied, simulation experiments comprise two stages: stage 1 for determining the length of the initial transient period and stage 2 during which the steady-state behavior is simulated and analyzed, as illustrated in Figure 3.

3.1 Detecting the Length of the Initial Transient Period

The sequential procedure presented here is based on a stationarity test proposed by Schruben et al. [1983]. It is used to test the hypothesis that a sufficient number of initial transient data has been (or has not been) discarded. As in any statistical test, the value of a chosen statistic calculated from the tested sequence of observations is

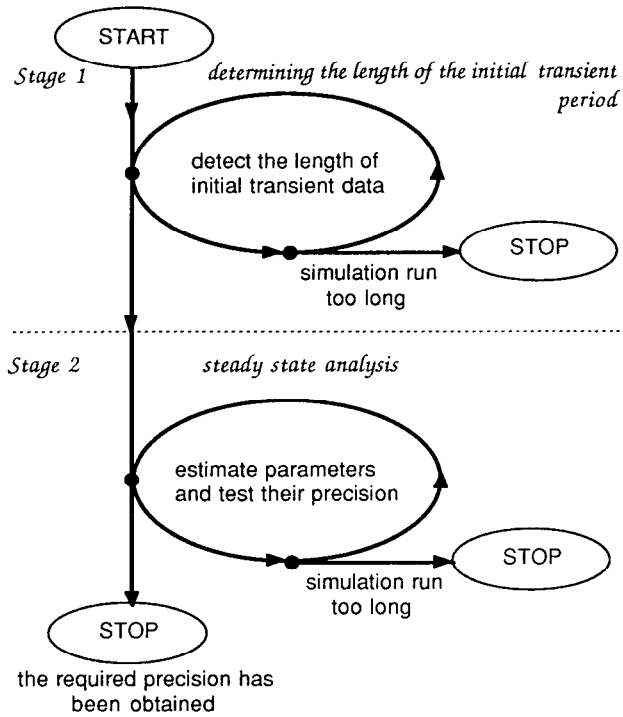

Figure 3. Two-stage sequential methods for data collection and analysis in steady-state simulation.

compared with the corresponding value from a standard sequence, and the decision about rejection or acceptance of the hypothesis is taken at an assumed significance level α_t, $0 < \alpha_t < 1$. The significance level can be regarded as the probability of erroneously rejecting the hypothesis that the tested process is stationary. To get a first approximation for the truncation point, n_o^*, we can use one of the heuristic rules R1–R10 presented in Section 2. For example, in simulation studies of data communication protocols [Pawlikowski and Asgarkhani 1988] rule R5 was applied (assuming $k = 25$). A flow chart of the procedure is given in Figure 4.

The problem encountered during testing a sequence of n_t observations for stationarity is that the steady-state estimator for the variance $\sigma^2[\bar{X}(n)]$, and the number κ of degrees of freedom for its χ^2 distribution, has to be known earlier than we know that the process has entered the stationary region. To get a robust estimate of that variance, the estimation should be done using only a subsequence of the last n_v observations from the sequence of n_t observations tested for stationarity; that is, taking $n_t \geq \gamma_v n_v$, where γ_v ($\gamma_v \geq 2$) is the "safety" coefficient for the variance estimator to represent the steady state. Assuming larger value of γ_v, the last n_v of n_t observations are more likely to be from the stationary region, even if the truncation point of the initial transient period has been initially underestimated. Both the value of n_v and n_t should be selected after having taken into account the minimum sample sizes required by the method of variance analysis and the stationarity test. Heidelberger and Welch [1983] assumed $n_v \geq 100$, $\gamma_v = 2$. Having assumed such value of n_v, we should have at least $n_t = 200$ observations stored in a buffer for testing against nonstationarity, which is the size of the sample assumed by Schruben [1982]. Since the number of observations tested for stationarity should be larger if longer transient

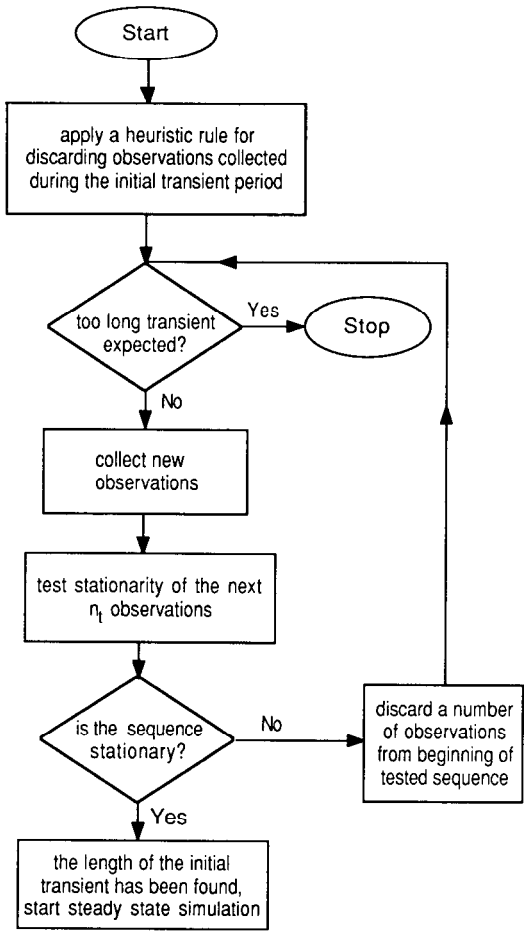

Figure 4. Sequential analysis of the length of the initial transient period.

periods are suspected, one can assume that

$$n_t = \max(\gamma_v n_v, \gamma n_o^*), \qquad (76)$$

where γn_o^* is the smallest length of one step in sequential testing for stationarity for a given n_o^*, $\gamma > 0$.

Thus, after having discarded n_o^* observations, the next n_t observations are collected and the last n_v observations are used to find $\hat{\sigma}^2[\bar{X}(n)]$ and κ. The variance and the degrees of freedom of its χ^2 distribution can be estimated using a few different methods, presented, for example, in Fishman [1973a, p. 289], Heidelberger and Welch [1981a], and Schruben [1982]. The method described by Heidelberger and Welch [1981a], based on the spectral analysis of observations, was used in simulation studies of data communication protocols [Asgarkhani and Pawlikowski 1989; Pawlikowski and Asgarkhani 1988] and is summarized in Appendix B.

Having estimated the variance, which by assumption represents steady state of the process, we can start testing the first n_t observations for stationarity. If the test accepts the hypothesis that the end of the initial transient has been detected correctly and the process has already entered its stationary region, this stage of analysis of the simulation output results is finished and the program may start to analyze confidence intervals; see Section 3.2. Otherwise transient effects have still been detected. Consequently, additional γn_o^* observations from the beginning of the test sequence are discarded, γn_o^* new observations are collected, the last n_v observations are again used to estimate $\sigma^2[\bar{X}(n_v)]$, and the first n_t observations in the stored sequence are tested for stationarity. This procedure is continued until the stationarity of the sequence of observations is confirmed or $n_{o,max}$ (the upper limit of the number of tested observations) is reached. As the longest acceptable length of the initial transient period, one can assume $n_{o,max} = 0.5n_{max}$, where n_{max} is the maximum length of the simulation run. If the transient phase extends beyond $n_{o,max}$ observations, either the allowed maximum length of the simulation run is too short or the system is unstable.

Summing, up, the procedure requires the following parameters:

n_{max} The maximum allowed length of the simulation run measured in the number of recorded observations (to be decided in advance. In simulation studies reported in Asgarkhani and Pawlikowski [1989] and Pawlikowski and Asgarkhani [1988] n_{max} was not less than 100,000);

$n_{o,max}$ The maximum allowed length of the initial transient period (the default value is $n_{o,max} = 0.5n_{max}$).

n_v The length of the sequence used for estimating the steady-state variance $\sigma^2[\bar{X}(n)]$ (the default value is $n_v = 100$).

n_t The length of the sequence tested for stationarity; see Equation (76).

α_t The significance level of the stationarity test ($0 < \alpha_t < 1$; the default value is $\alpha_t = 0.05$).

γ_v The "safety" coefficient for the estimator of variance $\sigma^2[\bar{X}(n)]$ to represent the steady state ($\gamma_v \geq 2$; the default value is $\gamma_v = 2$).

γ The "exchange" coefficient, determining the number of new observations included in each sequential test for stationarity ($\gamma > 0$; the default value is $\gamma = 0.5$).

The procedure can be described as follows:

procedure ***DetectInitial Transient***;
{*determine the length of the initial period applying the Schruben's test preceded by an heuristic rule of truncation*}

Step 1
Start the simulation run from the empty-and-idle state;
apply one of the heuristic deletion rules R1–R10 {*see Section 2*} to determine n_o^*;
{n_o^* *is the first approximation of the number of observations to be deleted*}
if (the initial transient period embraces more than $n_{o,max}$ observations) **then**
goto *Step 6*
else $n_o := n_o^*$; discard first n_o observations
endif;
if $\gamma_v n_v \geq \gamma n_o^*$ **then** $n_t := \gamma_v n_v$ **else** $n_t := \gamma n_o^*$
endif;
$\Delta_n := \gamma n_o^*$; {n_t *observations will be tested for stationarity,* Δ_n *old observations will be replaced by new ones*}
$\alpha_{t1} := \alpha_t$; {*the initial value of significance level of the test for stationarity; see Step 4*}

Step 2
If $n_o + n_t \leq n_{o,max}$ **then**
append Δ_n observations to the tested sequence;
{*some of these observations may have been already collected when the heuristic deletion rule was applied*}
goto *Step 3*
else {*if* $n_o + n_t > n_{o,max}$}
goto *Step 6*
endif;

Step 3
Determine the variance $\hat{\sigma}^2[\bar{X}(n_v)]$ and κ, the degrees of freedom of the variance distribution, using the last n_v collected observations starting from the observation ($n_o + n_t - n_v + 1$);
{*for example, apply the procedure Spectral-Variance Analysis described in Appendix B, assuming* $x_s = x_{n_o+n_t-n_v+s}$, *for* $s = 1, 2, \ldots, n_v$}

Step 4 {*the test for stationarity* [*Schruben et al. 1983, p. 1173*]}
Take all n_t observations, starting from the observation ($n_o + 1$), and calculate the test statistic

$$T = \frac{\sqrt{45}}{n_t^{1.5} n_v^{0.5} \hat{\sigma}[\bar{X}(n_v)]} \sum_{k=1}^{n_t} k\left(1 - \frac{k}{n_t}\right)[\bar{X}(n_t) - \bar{X}(k)],$$

$$\text{where } \bar{X}(i) = \sum_{j=n_o+1}^{n_o+i} \frac{x_j}{i};$$

if (a negative bias of the mean $\bar{X}(n_t)$ is suspected)
then goto *Step 5*
elsif (a positive bias is suspected) **then** $T := -T$
{*the reason for considering bias of an assumed sign is that one-sided tests are usually more powerful than their two-sided correspondents*}
else {*if a sign of the initial bias is difficult to predict, then prepare data for a two-sided test*}
$T := |T|$; $\alpha_{t1} := \alpha_t/2$
endif;

Step 5
if $T \leq t_{\kappa,1-\alpha_{t1}}$ **then**
{$t_{\kappa,1-\alpha_{t1}}$ *is the upper* $(1 - \alpha_{t1})$ *critical point from the t-distribution with* κ *degrees of freedom*}
write ('*the initial transient period is not longer than* n_o *observations*');
start sequential analysis of confidence intervals
{*call one of procedures presented in Section 3.2*}
else {*if* $T > t_{\kappa,1-\alpha_{t1}}$}
discard first γn_o^* observations from the tested sequence;
$n_o := n_o + \gamma n_o^*$; $\Delta_n := \gamma n_o^*$; **goto** *Step 2*
endif

Step 6 {*if the initial transient period embraces more than* $n_{o,max}$ *observations*}
stop the simulation run;
write ('*the initial transient period embraces more than* $n_{o,max}$ *observations, or the simulated process is unstable*')
end ***DetectInitialTransient***.

The effectiveness of this procedure strongly depends on the effectiveness of the variance estimator $\hat{\sigma}^2[\bar{X}(n)]$.

3.2 Sequential Testing for a Required Precision of Results

Both sequential procedures for stopping a simulation experiment that are presented here require the analyzed sequence of observations to be representative of steady state, so n_o observations representing the initial transient period have to have been discarded beforehand.

The first procedure is a modified version of the spectral method of analysis proposed by Heidelberger and Welch [1981a, 1981b, 1983]. Its simplified flow chart is given in Figure 5. As mentioned in Section 1, the method allows the reduction of the number of individual data items stored in memory during a simulation experiment by batching individual observations into batches of size $2m$, $m = 1, 2, \ldots$, and replacing them with the batch means. These means are stored in the buffer AnalysedSequence of size $2M$. Whenever $2M$ batch means over batches of m observations are recorded, they are consolidated into M means over batches of size $2m$. Subsequent observations are lumped into M successive batches of size $2m$. If more than M such new batches are needed, the rebatching procedure is repeated. Thus, the buffer AnalysedSequence can be implemented simply as a one-dimensional array of size $2M$. The accuracy of estimators is measured by the relative precision ϵ of confidence intervals, defined in Equation (12), and the simulation is stopped if $\epsilon \leq \epsilon_{max}$, where ϵ_{max} is the acceptable maximum value of the relative precision of the final results. The current values of ϵ are evaluated at consecutive checkpoints ω_k ($k = 1, 2, \ldots$; $\omega_k \leq n_{max}$); that is, each time $(\omega_{k+1} - \omega_k)$ new observations have been collected. To limit the number of possible checkpoints, one can assume that they are geometrically distributed; that is, for a given ω_1,

$$\omega_{k+1} = \min\{\lfloor \gamma_a(\omega_k - n_o) \rfloor + n_o, n_{max}\}, \quad (77)$$

where $k = 1, 2, \ldots$; $\lfloor x \rfloor$ denotes the greatest integer not greater than x and $\gamma_a > 1$. To avoid too large a distance between consecutive checkpoints, one can assume $\omega_1 = \max(2M, 2n_o)$ and a constant value for the step between checkpoints above a threshold point. Following a different reasoning, selection of the value of ω_1 satisfying the inequality

$$\lfloor 0.1(n_{max} - n_o) \rfloor + n_o \leq \omega_1 \leq \lfloor 0.2(n_{max} - n_o) \rfloor + n_o \quad (78)$$

was suggested in Heidelberger and Welch [1983]. The estimator of variance $\sigma^2[\bar{X}(n)]$ and the number κ of its degrees of freedom needed in analysis of confidence intervals are calculated here using the procedure SpectralVarianceAnalysis. It requires at least $n_v \geq 100$ batch means to be available (see Appendix B). This approach is described below by a pseudocode procedure that uses the following parameters:

n_o The number of discarded initial observations (determined by the procedure DetectInitialTransient).

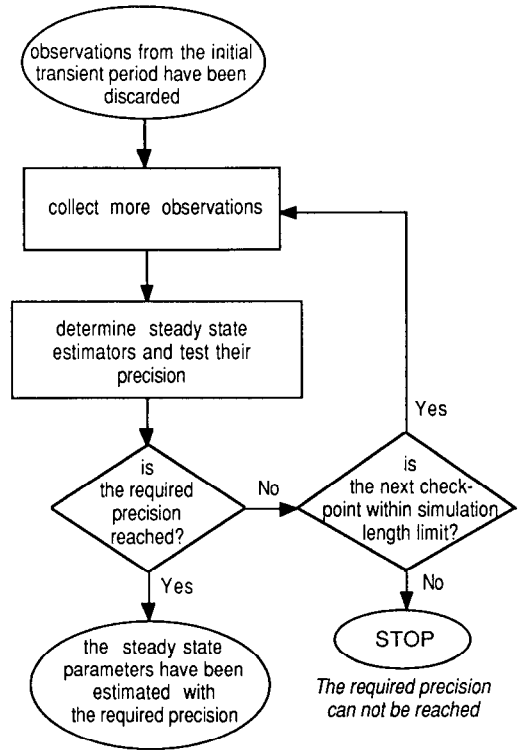

Figure 5. Flow chart for the spectral method.

n_{max} The maximum allowed length of the simulation run, measured in the number of recorded observations ($n_{max} \geq \max(3n_o, n_o + 2M)$; to be decided in advance).

n_v The length of the sequence used for estimating the variance $\sigma^2[\bar{X}(n)]$ ($n_v \geq 100$; the default value is $n_v = 100$).

γ_a The checkpoint incremental coefficient for sequential testing for accuracy ($\gamma_a > 1$; the default value is $\gamma_a = 1.5$).

$(1 - \alpha)$ The assumed confidence level of the final results ($0 < \alpha < 1$; the default value is $\alpha = 0.05$).

ϵ_{max} The maximum acceptable value of the relative precision of confidence intervals ($0 < \epsilon_{max} < 1$; the default value is: $\epsilon_{max} = 0.05$).

procedure ***SpectralAnalysis***;
{*sequential analysis of simulation output data based on spectral analysis of the series of collected observations*}
const $M = 100$;
{*the default value of the minimum number of data points for the analysis of confidence intervals*; $M \geq n_v \geq 100$; *see Section 3.1*}

procedure ***Batching***;
{*preparation of secondary output data; transformation of individual observations into the sequence of no more than 2M batch means of sequentially increasing batch size*}
begin {*calculate the batch mean* $\bar{X}_j(m)$, *and store it as the jth data item in the buffer AnalysedSequence*}
$\bar{X}_j(m) := sum/m$;
if $j = 2M$ **then** {*consolidate 2M means of batches of size m into M means of batches of size 2m*}
for $s := 1$ **to** M **do**
$\bar{X}_s(2m) := 0.5(\bar{X}_{2s-1}(m) + \bar{X}_{2s}(m))$
enddo
$m := 2m$; $j := M$
endif
$j := j + 1$; sum := 0
{*start to calculate the next batch mean*}
end ***Batching***;

procedure ***Estimation***;
{*sequentially calculate estimates and test their precision until the required precision is reached*}
begin
find the estimator $\hat{\sigma}^2[\bar{X}(n_v)]$ for the sequence $\bar{X}_{j-n_v+1}(m), \ldots, \bar{X}_j(m)$, the last n_v batch means stored in the buffer AnalysedSequence, and determine κ, the degrees of freedom of the distribution of $\hat{\sigma}^2[\bar{X}(n_v)]$;
{*apply the procedure SpectralVarianceAnalysis; see Appendix B*}
calculate the relative half width of the confidence interval at the confidence level $(1 - \alpha)$ for the current checkpoint ω_k:

$$\epsilon = t_{\kappa,1-\alpha/2}\hat{\sigma}[\bar{X}(n_v)]/\bar{X}(jm),$$

where $\bar{X}(jm) = \sum_{s=1}^{j} \bar{X}_s(m)/j$

is the current value of the estimated mean after jm observations, and $t_{\kappa,1-\alpha/2}$ is the upper $(1 - \alpha/2)$ critical point of the t-distribution with κ degrees of freedom;
{*test conditions of stopping the simulation run*}
if $(\epsilon \leq \epsilon_{max})$ **then**
{*print the final results and stop the simulation*}
StopSimulation := **true**
else {*the required precision has not been reached yet; determine the next,* $(k + 1)$*st, checkpoint*}
$k := k + 1$;
$\omega_k := \min(\lfloor \gamma_a(\omega_{k-1} - n_o) \rfloor + n_o, n_{max})$
endif
end ***Estimation***;

begin {*main procedure*}
$m := 1$; {*the initial batch size*}
$k := 1$;
{*the initial checkpoint is after* $\omega_k = \omega_1$ *observations*}
$\omega_o := n_o$; $i := 1$;
{*having discarded* n_o *observations, collect next observations starting from the observation* $(\omega_o + i)$}
$\omega_1 := \max(2M, 2n_o)$;
{*the default location of the first checkpoint*}
sum := 0; $j := 1$;
{*start calculating the 1st batch mean*}
StopSimulation := **false**;
{*a condition of stopping the simulation has not been met yet*}
while (**not** StopSimulation) **do**
{*collect and process new* $(\omega_k - \omega_{k-1})$ *observations*}
get the observation x_{ω_o+i};
sum := sum + x_{ω_o+i};
if $(i$ **mod** $m = 0)$ **then** ***Batching*** **endif**;
if $(i = \omega_k)$ **then** ***Estimation*** **endif**;
if (**not** StopSimulation) **then**
$i := i + 1$;
if $(i > n_{max} - n_o)$ **then**
write ('*the required accuracy cannot be reached; either restart the simulation*

```
            using a new seed (seeds) of random
            number generator, or increase n_max, or
            increase α, or increase ε_max');
          {for obtaining a new sequence of pseu-
           dorandom numbers, independent from
           the previous one, assume the last num-
           ber from the previous sequence as the
           first number of the new sequence}
          StopSimulation := true
        endif
      endif
    enddo;
    write ('the required precison of results has
      been obtained having collected', n_o + jm,
      'observations; the final relative precision:',
      100ε%, 'the final (1 − α)100% confidence
      interval:', X̄(jm)[1 ± ε])
end SpectralAnalysis.
```

The next sequential procedure for stopping simulation experiments when the required precision of results is reached is based on the method of nonoverlapping batch means. Its simplified flow chart is given in Figure 6 (only the main loops of computations are depicted). As in the case of the procedure SpectralAnalysis, the initial nonstationary sequence of n_o observations should have been discarded earlier, applying, for example, the procedure DetectInitialTransient. For weakening serial correlations of analyzed output data, individual observations are replaced here by the less correlated means of their batches. Thus, the problem of direct analysis of confidence intervals from correlated observations is replaced by the problem of determining the batch size m^*, such that batch means are (almost) uncorrelated at a given level of significance. Generally, in a sequence of correlated data the autocorrelation coefficients of lag k, $k = 1, 2, \ldots,$ [see Equations (21) and (22)] are not necessarily decreasing as the lag increases, although all autocorrelation coefficients are zero if the sequence contains uncorrelated observations. For this reason, we follow the test proposed by Adam [1983]: A given batch size can be accepted as the batch size for approximately uncorrelated batch means if all L autocorrelation coefficients of lag k ($k = 1, 2, \ldots, L$) are statistically negligible at a given significance level β_k, $0 < \beta_k < 1$. The analytical problems encountered during estimation of the autocorrelation coefficients suggest that the number of considered lags should be limited to $L = 0.1\ k_{bo}$, where k_{bo} is the number of batch means tested for autocorrelation; see comments in Section 1. The autocorrelation coefficients can be better estimated by the so-called jackknife estimators [Miller 1974], which are usually less biased than the ordinary estimators defined by Equations (21) and (22). A jackknife estimator of autocorrelation coefficient of lag k for a sequence of batch means of size m is calculated from the following formula:

$$\hat{\hat{r}}(k, m) = 2\hat{r}(k, m) - \frac{\hat{r}'(k, m) + \hat{r}''(k, m)}{2}, \tag{79}$$

where the estimators on the right-hand-side are calculated like ordinary estimators of autocorrelation coefficients [see Equations (21) and (22)], except that $\hat{r}(k, m)$ is the estimator over all k_{bo} batch means, whereas $\hat{r}'(k, m)$ and $\hat{r}''(k, m)$ are estimators over the first and the second half of the analyzed sequence of k_{bo} batch means, respectively.

Let us note the following:

(i) To get acceptable estimators of the autocorrelation coefficients, at least 50 batch means should be available [Box and Jenkins 1970, p. 33]; thus, in the case of jackknife estimators one should assume $k_{bo} \geq 100$.

(ii) To ensure approximate normality of batch means, the size of considered batches should not be less than 50 [Adams 1983].

(iii) To get an acceptable overall significance level β when testing the value of L autocorrelation coefficients of lag k ($k = 1, 2, \ldots, L$), each at the significance level β_k, we have to assume

$$\beta < \sum_{k=1}^{L} \beta_k, \tag{80}$$

hence in practice L should not be too large. This restriction is irrelevant if the autocorrelation coefficients decrease monotonically with the value of the lag, since then only $\hat{\hat{r}}(1, m)$ has to be considered.

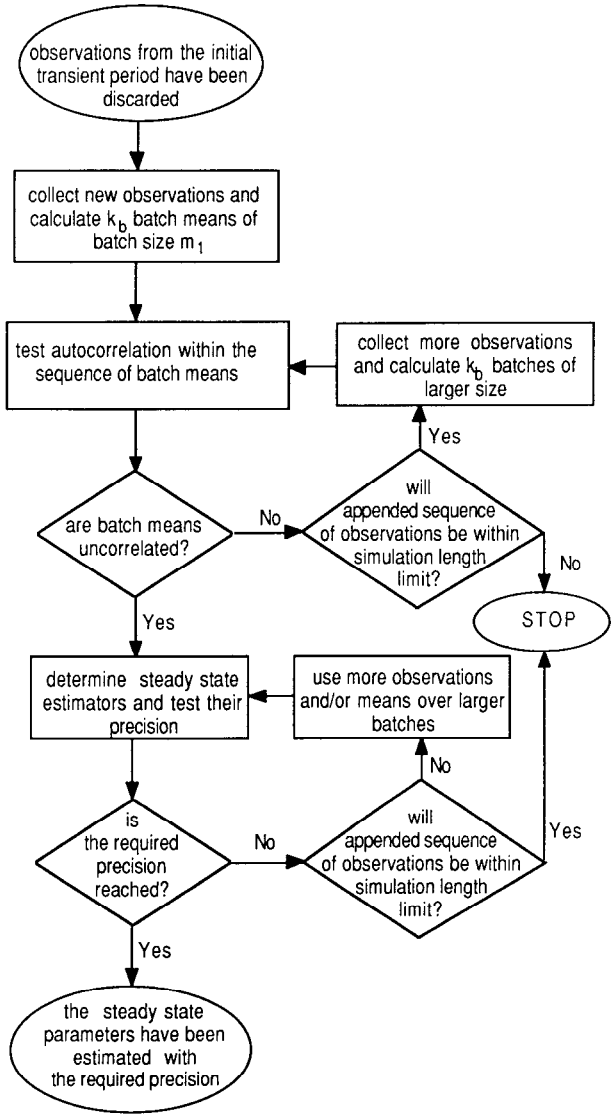

Figure 6. Flow chart for the method of batch means.

To avoid wastefully collecting an excessive number of observations, especially when testing batch sizes, the procedure BatchMeansAnalysis uses two buffers for storing batch means: A buffer called ReferenceSequence is used to store the batch means $\bar{X}(m_o)$, $\bar{X}_2(m_o)$, . . . of a batch size m_o, and a buffer AnalysedSequence is used for storing an assumed number k_{bo} of batch means over batches of size $m_s = sm_o$ ($s = 1, 2, \ldots$), formed from the batch means kept in the ReferenceSequence. Thus, since the number of data items collected in the ReferenceSequence grows in time during a simulation run, a linked list of batch means seems to be a proper data structure for this buffer. The number of data items in the AnalysedSequence is limited to k_{bo}, so it can be implemented as an ordinary one-dimensional array. By selecting m_o

properly, we can secure a sequential increase of tested batch sizes slower than in the batching schemes proposed in Adam [1983], Fishman [1973a], Law and Carson [1979] to reduce the resultant simulation run length. For neutralizing the observed randomness of the estimators of correlation coefficients, $m^* = m_s$ is selected as the final batch size of weakly correlated means iff the hypothesis of all zero autocorrelation coefficients is accepted in two succesive tests, both for the batch size m_{s-1} and m_s.

Having selected the batch size m^*, one can sequentially analyze the accuracy of results by calculating confidence intervals from Equations (18) and (19), which are valid for independent and identically distributed batch means. The sequence of batch means for batch size m^* kept in the buffer ReferenceSequence can be sequentially appended by new batch means if more observations are needed to improve the accuracy of results. As mentioned in Section 1, Schmeiser [1982] showed that using 10 to 30 batch means over longer batches can give more accurate results (and a better coverage of the estimators) than using more batch means over smaller batches. Following these recommendations, when the accuracy test of the estimator from k_{be} batch means stored in the ReferenceSequence fails, these k_{be} batch means are used to form $k_{bo} = 30$ batch means in the AnalysedSequence buffer for the additional accuracy test. Such a test is done before a new batch mean is appended to the ReferenceSequence.

The whole method can be summarized by the following pseudocode procedure, which requires these parameters:

n_o The number of discarded initial observations (determined by the procedure DetectInitialTransient).

n_{max} The maximum allowed length of the simulation run, measured as the number of recorded observations ($n_{max} \geq n_o + m_o k_{bo}$; to be determined in advance).

$1 - \alpha$ The assumed confidence level of the final results ($0 < \alpha < 1$; the default value is $\alpha = 0.05$).

ϵ_{max} The maximum acceptable value of the relative precision of confidence intervals ($0 < \epsilon_{max} < 1$; the default value is $\epsilon_{max} = 0.05$).

The initial values of other parameters are given in the procedure.

```
procedure BatchMeansAnalysis;
{sequential analysis of simulation output data
 based on analysis of uncorrelated means of non-
 overlapping batches of observations}
const m_o = 50;
  {the default value of the batch size for means
   stored in the ReferenceSequence}
k_bo = 100;
  {the default value of the number of batch means
   stored in the AnalysedSequence}

procedure DetermineBatchSize;
{determine the batch size for approximately un-
 correlated batch means}
begin
  s := 1; j := 1; sum := 0;
    {these are the initial values of parameters for
     determining X̄_j(m_o), the jth batch mean of
     size m_o, at the sth sequential step}
  while ((not StopSimulation) and (not Un-
      correlated)) do
    {(i − 1) observations have been already re-
     corded}
    for v := i to i + k_bo m_o − 1 do
      {collect new k_bo m_o observations for the next
       k_bo batch means of size m_o to be stored in
       the ReferenceSequence}
      get the observation x_{n_o+v};
        {the initial n_o observations have been
         discarded}
      sum := sum + x_{n_o+v};
      if (v mod m_o = 0) then
        {calculate the batch mean X̄_j(m_o); the
         jth data item in the Reference-
         Sequence}
        X̄_j(m_o) = sum/m_o; sum := 0; j := j + 1
      endif  {the next batch mean has been
              determined}
    enddo  {the next k_bo batch means have been
            determined}
    for v := 1 to k_bo do
      {consolidate sk_bo batch means X̄_1(m_o),
       X̄_2(m_o), ..., from the ReferenceSequence
       into k_bo batch means Ȳ_1(sm_o), Ȳ_2(sm_o),
       ..., in the AnalysedSequence, which will
       be tested for autocorrelation}

      Ȳ_v(sm_o) := Σ_{r=1}^{s} X̄_{(v−1)s+r}(m_o)/s
```

```
    enddo
    TestCorrelation;
    i := i + k_bo m_o;
      {n_{o+i-1} observations have been collected}
    if (not Uncorrelated) then
      {re-initialize the counter variable for the
       next autocorrelation test}
      s := s + 1
    endif
    if (not Uncorrelated) and (n_o + i + k_bo m_o
        - 1 > n_max) then
      {stop selecting the batch size if the next
       test requires more than n_max observations}
      write ('the batch size of uncorrelated
          batch means cannot be determined;
          increase n_max or β');
          {β is the significance level of tests
          against autocorrelation used in the
          procedure TestCorrelation; see Eq.
          (80)}
      StopSimulation := true
    endif
  enddo  {of search for the batch size of uncor-
          related batch means}
end DetermineBatchSize;

procedure TestCorrelation;
{test significance of autocorrelations between
 batch means for a given batch size}
const L = 10;
  {the default number of autocorrelation coeffi-
   cients; if autocorrelation coefficients mono-
   tonically decrease with the value of lag take
   L = 1, else L := 0.1 k_bo}
β = 0.1;  {the default value of the overall signif-
           icance level of L tests against autocor-
           relation}
begin
  Correlation := 0;  {the initial value for testing
                      L correlation coefficients}
  for k := 1 to L do
    {test whether all L autocorrelation coeffi-
     cients are statistically negligible each at the
     β_k significance level}
    calculate the jackknife estimator r̂(k, sm_o)
    of the autocorrelation coefficient of lag k
    for the sequence Ȳ_1(sm_o), Ȳ_2(sm_o), ...,
    Ȳ_{k_bo}(sm_o) stored in the Analyzed-
    Sequence; {apply Eq. (79) for m = sm_o}
    β_k := β/L;
    σ̂²[r̂(k, sm_o)] :=
      if (k = 1) then (1/k_bo)
        else ([1 + 2 Σ_{u=1}^{k-1} r̂²(u, sm_o)]/k_bo)
      endif
      {the variance of the autocorrelation coef-
       ficient of lag k [Bartlett 1946]}
    if ( | r̂(k, sm_o) | < z_{1-β_k/2} σ̂[r̂(k, sm_o)])
      {z_{1-β_k/2} is the upper (1 - β_k/2) critical point
       of the standard normal distribution}
      then r̂(k, sm_o) := 0
    endif  {the lag k autocorrelation is statisti-
            cally negligible at the confidence
            level 1 - β_k}
    Correlation := Correlation + | r̂(k, sm_o) |;
  enddo
  if ((Correlation = 0) and AcceptableSize)
then
    {accept the current batch size since the pre-
     vious batch size has already given negligible
     correlations}
    m* := sm_o;
    Uncorrelated := true
    {the batch size m* of uncorrelated batch
     means has been selected}
  elsif (Correlation = 0) then
    {start collecting next observations for testing
     larger batch size}
    AcceptableSize := true
    {autocorrelations for the current batch size
     are negligible but they were not negligible
     for the previous batch size, thus the next
     batch size should be considered}
  endif
end TestCorrelation;

procedure Estimation;
{sequentially calculate estimates and test their
 precision until the required precision is reached}
begin
  calculate the mean X̄(k_be, m), its variance
  σ̂²[X̄(k_be, m)], and the relative half width ε of
  the confidence intervals at the confidence
  level (1 - α):

  ε := t_{k_be-1,1-α/2} σ̂[X̄(k_be, m)]/X̄(k_be, m),

  using the whole sequence of k_be data items
  stored in the ReferenceSequence:
  {apply Eqs. (17)-(19), for k_b := k_be, m := m*
   and X̄_i(m) := X̄_i(m*)}
  if ((ε > ε_max) and (k_be mod 30 = 0)) then
    {the additional test of accuracy of estimators
     after consolidation of k_be batch means from
     the ReferenceSequence into k_bo = 30 means
     of longer batches stored in the Analysed-
     Sequence; see the recommendations in
     Schmeiser [1982]}
    m := k_be/30;
```

```
    for j := 1 to 30 do
      Ȳ_j(mm*) := Σ_{r=1}^{m} X̄_{(j-1)m+r}(m*)/m;
      calculate the mean X̿(k_bo, mm*), its variance σ̂²[X̿(k_bo, mm*)], and the relative
      half width ε of the confidence interval at the confidence level (1 − α):
      ε :=
        t_{k_bo−1,1−α/2} σ̂[X̿(k_bo, mm*)]/X̿(k_bo, mm*),
      from the sequence of 30 data items stored in the AnalyzedSequence;
      {apply Eqs. (17)–(19), for k_bo := 30, m := mm*, X̄_i(m) := Ȳ_i(mm*)}
    enddo
  endif
  if (ε ≤ ε_max)  {the required precision has been reached}
    then write ('the required precision of results has been obtained after', n_o + k_be m*, 'observations recorded');
    StopSimulation := true
  elsif (n_o + (k_be + 1)m* ≤ n_max) then
    {k_be batch means of size m* have been used to estimate the mean μ_x, but the required precision has not been reached yet; collect the next batch of m* observations and store their mean in the ReferenceSequence}
    sum := 0;
    for j := 1 to m* do
      get the observation x_{k_be m*+j};
      sum := sum + x_{k_be m*+j}
    enddo
    k_be := k_be + 1;
    X̄_{k_be}(m*) := sum/m*
  else {the required precision has not been reached; too short the simulation run assumed}
    write ('the required accuracy cannot be reached; either restart the simulation using a new seed (seeds) of random number generator, or increase n_max, or increase α, or increase ε_max');
      {for obtaining a new sequence of pseudorandom numbers, independent from the previous one, assume the last number from the previous sequence as the first number of the new sequence}
      StopSimulation := true
  endif
end Estimation;

begin {main procedure}
  StopSimulation := false;
    {a condition of stopping the simulation has not been met yet}
  AcceptableSize := false;
    {a batch size for uncorrelated batch means has not been found yet}
  Uncorrelated := false;
    {the batch size for uncorrelated batch means has not been determined yet}
  ε := 1 − ε_max;  {the initial precision ε > ε_max}
  i := 1;  {having discarded n_o observations, collect the next observations starting from the observation (n_o + 1)st}
  DetermineBatchSize;
    {if the batch size m* has been selected then the ReferenceSequence contains sk_bo batch means of size m*/s and the AnalysedSequence contains k_bo batch means of size m*}
    k_be := k_bo;
    {the size of the ReferenceSequence at the beginning of estimation}
  for j := 1 to k_be do
    {prepare data for estimation; consolidate sk_bo batch means of size m_o = m*/s in the ReferenceSequence into k_be = k_bo batch means of size m*}
    X̄_j(m*) := Σ_{r=1}^{s} X̄_{(j-1)s+r}(m_o)/s
  enddo
  while (Uncorrelated and (not StopSimulation))
  do
    {sequentially test the precision of the estimators until the required precision or the maximum length of simulation run is reached}
    Estimation;
  enddo
  write ('the final relative precision:', 100ε%, 'the final 100(1 − α)% confidence interval: ', X̿(k_be, m*)[1 + ε]);
end BatchMeansAnalysis.
```

In practice, the last procedure often uses a sequential search for the batch size of uncorrelated batch means rather than the sequential stopping rule for the simulation run (sequential testing the precision of estimates). This happens because usually many more batch means have to be tested

during the first stage (testing against autocorrelation) than during the second stage (testing the precision of the results). The number of observations recorded, $k_{bo}m^*$, when the mean and the width of confidence intervals are to be calculated for the first time is usually much larger than is required for obtaining the required level of accuracy [Schmeiser 1982]. From this point of view, the spectral method of analysis is more thrifty. On the other hand, the method of batch means uses a very simple estimator of the variance when the width of confidence intervals is analyzed. Exhaustive comparative studies of both procedures have not been performed yet, but the reported results indicate that the spectral method is usually more efficient, both in the sense of the coverage recorded in reference experiments and in the sense of the final simulation run lengths for obtaining the required accuracy of the results, at least in investigated classes of processes [Pawlikowski and Asgarkhani 1988].

4. SUMMARY AND GENERAL COMMENTS

We have discussed in detail methods for dealing with the main phenomena encountered in steady-state simulation of queuing processes: the inherent initial nonstationarity and the permanent autocorrelation of collected observations. The emphasis is on methods for the sequential analysis of simulation output data, bearing in mind their possible implementation in user-friendly simulation packages, which would produce results automatically. In such a context, methods of analysis based on single simulation runs seem to be more attractive than the methods of independent replications.

This survey has been limited to analysis of a single measure of performance, namely, the point and interval estimates of the sample mean. Nevertheless, this methodology can easily be modified to estimate other performance measures, as long as their estimators are based on cumulative statistics. For example, sequential analysis of simulation output data can easily be applied to estimate the variance, or the probability that the analyzed variate lies in some fixed interval [Welch 1983, Sec. 6.3]. It cannot be applied directly to estimate quantiles, since their estimators require that sequences of collected observations be presorted and as such are not amenable to cumulative statistics calculated while the simulation is in progress. Estimation of quantiles by using the method of spectral analysis is discussed in Heidelberger and Lewis [1984]; the same application of the regenerative method is considered in Iglehart [1976] and Seila [1982a]. Computational problems associated with quantile estimation are discussed in Jain and Chlamtac [1985] and Raatikainen [1987, 1988].

The methodology of simultaneous analysis of more than one measure of performance during one simulation experiment is much less advanced than the methodology of univariate estimation discussed in this paper. The main theoretical problems of multivariate estimation and the methods that can be used for determining the confidence regions in such multiresponse simulation experiments are discussed in Charnes and Kelton [1988], Chen and Seila [1987], Friedman [1984], Law and Kelton [1982a, p. 308], and Seila [1982b, 1983]. The initialization bias in multiresponse simulation can be controlled by applying a statistical test proposed by Schruben [1981]. Various variance reduction techniques that are applicable in this kind of simulation experiments are discussed in Bauer et al. [1987], Rubinstein and Marcus [1985], Venkatraman and Wilson [1986], and Yang and Nelson [1988].

Specific statistical problems accompanying comparative simulation studies of alternative systems are surveyed in Law and Kelton [1982a, Chap. 9]; see also Balmer [1987], Clark [1986], Friedman and Friedman [1986], and Goldsman [1983, 1986]. (Note that in these applications the bias of estimators is not important, as long as the estimators are equally biased.) Stochastic simulation is also used for sensitivity analysis of performance measures and optimization of various systems and processes under uncertainty. Special methods for estimating the gradient of a performance measure with respect to a selected input

parameter (e.g., the changes of the mean delay of messages in a communication network with respect to the mean message length) during a single simulation run are discussed in [Ho 1987; Ho and Cao 1983; Ho et al. 1983; Reiman and Weiss 1986; Rubinstein 1989; Suri and Zazanis 1988].

This paper is restricted to the statistical analysis of simulation output data, but practitioners are aware that that is not the only problem that must be overcome in obtaining useful results. Before observations are collected, the processes for which performance is to be investigated have to be properly modeled, and each model should be validated and verified to make the simulation experiments credible. Various aspects of the validation and verification of simulation models are discussed in [Banks and Carson 1984; Bulgren 1982; Law and Kelton 1982a; Sargent 1986; Shannon 1981; Velayas and Levary 1987].

Having presented methods of data collection and analysis that are used in stochastic steady-state simulation, one can conclude that no definite conclusions can be made about their applicability. The need for more exhaustive comparative studies expressed by Schriber and Andrews [1981] still seems to be a live issue.

The length of simulation runs remains a critical issue, especially in the case of simulation studies of complex systems, which often can be performed only if a technique for speeding up the process of data collection is applied. In this context an important role could be played by the variance reduction techniques, which, by reducing the variance, narrow confidence intervals and consequently reduce the number of steps needed by sequential procedures for reaching the required accuracy of results. But, as was mentioned in the Introduction, practical implementations of variance reduction techniques have been reported infrequently. Among the recently published ones are those in Izydorczyk et al. [1984] and Walrand [1988]. Another area in which the duration of simulation can be very critical is the analysis of performance measures that depend on events occurring infrequently. Some techniques proposed for speeding up rare events simulation are discussed in Cottrell [1983], Parekh and Walrand [1989], and Shahabuddin et al. [1988].

Performance evaluation studies based on simulation experiments can be accelerated by the decomposition of analyzed systems into subsystems, which are modeled separately, applying both simulation and analytical models, mutually interacting if necessary. The efficiency of such an approach was studied extensively in Blum et al. [1985] in the context of queuing network models of computer systems. A survey of hybrid techniques can be found in Shanthikumar and Sargent [1983]; see also Frost et al. [1988] and Kurose and Mouftah [1988]. The time needed for simulation studies can also be significantly reduced in multiprocessor systems. In such an environment individual processors can be used for running independent replications of one simulation experiment, or, in a more sophisticated solution, logical processes occurring during one simulation run can be executed in parallel by different processors. The main problem encountered in the latter case is the synchronization among processes run on different processors. There exists a danger that the contribution of large synchronization overheads will actually slow the simulation experiment. Various techniques of distributed and parallel simulation are discussed in Chandy and Misra [1981], Fujimoto [1988], Misra [1986], Nicol [1988a, 1988b], Reynolds [1988], Unger [1988], and Wagner and Lazowsa [1989]. See also Baik and Zeigler [1985], DeCegama [1987], Krishnamurthy et al. [1985], Kumar [1986], and Kurose and Mouftah [1988]. Specific problems relating to the statistical analysis of output data that accompany distributed simulation are discussed in Heidelberger [1986].

The diversity of methodological approaches and the complexity of some statistical techniques used in simulation output analysis has motivated attempts to develop expert systems that give advice on selecting the best simulation methodology for accomplishing the required statistical computations; see, for example, Haddock [1987], Hahn [1985], Hand [1985], Nielsen [1986], and O'Keefe [1986]. One of the first

statistical expert systems specifically oriented toward steady-state simulation is discussed in Mellichamp and Park [1989]. It is hoped that further developments in the area of expert systems and applied statistics will make it possible to design fully automated, knowledge-based simulation packages.

APPENDIX A. ELEMENTARY STATISTICAL CONCEPTS

Appendix A contains elementary statistical concepts needed to understand the problems discussed in this paper. A deeper and more detailed discussion of the subject can be found in the numerous textbooks such as [Trivedi 1982] on applied probability theory and statistics.

Stochastic simulation involves experiments that mimic various events by means of numbers generated by (pseudo)-random numbers generators. Therefore, it is natural that the analysis of the output data should be based on the methods of statistical inference. These methods are applied to estimate the characteristics and/or parameters of simulated processes. Accepting the random nature of simulation output data, one should consider any sequence $x_1, x_2, \ldots, x_n$ of observations collected during such experiments as the sequence of realizations of random variables $X_1, X_2, \ldots, X_n$ sampled from a stochastic process $X(t)$ at instants of time $t = t_i$, $i = 1, 2, \ldots, n$. If the process $X(t)$ is stationary, which means that its stochastic characteristics are time invariant, then the random variables $X_1, X_2, \ldots, X_n$ have the same (but unknown) probability distribution. In the context of this paper, the process is stationary if it enters its steady-state region, and steady-state simulation is concerned with modeling and analysis of such processes only.

Any function $f(X_1, X_2, \ldots, X_n)$ of random variables $X_1, X_2, \ldots, X_n$ is called a *statistic*. A statistic $\hat{\theta}(X_1, X_2, \ldots, X_n)$ used for estimating a parameter θ of the distribution of the analyzed process $X(t)$ is called an *estimator* of θ. The number of independent random variables used in the definition of a statistic is called its *number of degrees of freedom* or, simply, its *degrees of freedom*. The value $\hat{\theta}(x_1, x_2, \ldots, x_n)$, that is, the value of function $\hat{\theta}(X_1, X_2, \ldots, X_n)$ for a given sequence of observations $x_1, x_2, \ldots, x_n$ is known as an *estimate* of θ. It means that estimators, as functions of random variables, are random variables themselves, and estimates are simply realizations of these random variables.

A good estimator should be *unbiased*; see [Trivedi 1982] for other desired properties of estimators. If the estimator $\hat{\theta}^* = \hat{\theta}^*(X_1, X_2, \ldots, X_n)$ of the parameter θ is unbiased, then its statistical average

$$\mathrm{E}[\hat{\theta}^*(X_1, X_2, \ldots, X_n)] = \theta. \quad \text{(A1)}$$

For example, it can be shown that the estimator $\bar{X}(n)$ given in Equation (1) is an unbiased estimator of the average μ_x of a stationary process $X(t)$, since $\mathrm{E}[\bar{X}(n)] = \mu_x$. Similarly, one can show that if random variables $X_1, X_2, \ldots, X_n$ are independent and identically distributed, then the unbiased estimator of the variance $\sigma^2[X]$ of the stationary process $X(t)$ is

$$\hat{\sigma}^2[X] = \sum_{i=1}^{n} \frac{\{X_i - \bar{X}(n)\}^2}{n-1}, \quad \text{(A2)}$$

while the unbiased estimator of the variance of the estimator $\bar{X}(n)$ is

$$\hat{\sigma}^2[\bar{X}(n)] = \frac{\hat{\sigma}^2[X]}{n}; \quad \text{(A3)}$$

see Equation (4).

Theoretical justification for using $\bar{X}(n)$ as the estimator of μ_x is provided by the *law of large numbers*, which states that if observations $x_1, x_2, \ldots, x_n$ are realizations of independent random variables, taken from a stationary process $X(t)$, then

$$\lim_{n\to\infty} \mathrm{P}[|\bar{X}(n) - \mu_x| \le \epsilon] = 1, \quad \text{(A4)}$$

for any $\epsilon > 0$.

To construct confidence intervals for $\bar{X}(n)$, we have to assume a probability distribution for $\bar{X}(n)$. If $\bar{X}(n)$ is the weighted sum of independent and normally distributed random variables X_i, $i = 1, 2, \ldots, n$, with unknown variance $\sigma^2[X]$, then it can be proved that the statistic

$$T_{n-1} = \frac{\bar{X}(n) - \mu_x}{\hat{\sigma}[\bar{X}(n)]} \quad \text{(A5)}$$

has the standard (Student) t-distribution with $(n-1)$ degrees of freedom. Using this distribution, one can easily find the value $t = t_{n-1,1-\alpha/2}$ for it, given α, $0 < \alpha < 1$,

$$P(-t_{n-1,1-\alpha/2} \leq T_{n-1} \leq t_{n-1,1-\alpha/2}) = 1 - \alpha. \quad \text{(A6)}$$

Consequently, the confidence interval of μ_x at the $(1-\alpha)$ confidence level is given by

$$P(\bar{X}(n) - t_{n-1,1-\alpha/2}\hat{\sigma}[\bar{X}(n)] \leq \mu_x \leq \bar{X}(n) + t_{n-1,1-\alpha/2}\hat{\sigma}[\bar{X}(n)]) = 1 - \alpha; \quad \text{(A7)}$$

see Figure 1. The limiting t-distribution for $n \to \infty$ is the standard normal distribution described by the density function

$$p_z(z) = \frac{1}{\sqrt{2\pi}} \exp(-z^2). \quad \text{(A8)}$$

In practice, the approximation of a t-distribution by the normal distribution is already acceptable for samples of $n > 30$ observations. In such cases, one can easily find the value $z = z_{1-\alpha/2}$ from the standard normal distribution for which

$$P(\bar{X}(n) - z_{1-\alpha/2}\hat{\sigma}[\bar{X}(n)] \leq \mu_x \leq \bar{X}(n) + z_{1-\alpha/2}\hat{\sigma}[\bar{X}(n)]) = 1 - \alpha. \quad \text{(A9)}$$

which is a good approximation for the confidence interval of μ_x at the $(1-\alpha)$ confidence level if $n > 30$. In the case of multivariate estimation, confidence intervals are replaced by confidence regions (rectangles, etc.).

A justification of the assumption that the estimator $\bar{X}(n)$ is a normal random variable even if observations $x_1, x_2, \ldots, x_n$ are not is given by the *central limit theorem.* This theorem says that the sums of independent and identically distributed random variables tend to be normally distributed, even though the components are not, if $n \to \infty$. From this theorem, one can show that the limit distribution (for $n \to \infty$) of the random variable

$$Z = \frac{\bar{X}(n) - \mu_x}{\sigma[X]/\sqrt{n}} \quad \text{(A10)}$$

is given by Equation (A8). Note that it requires the variance $\sigma^2[X]$ of the random variables $X_1, X_2, \ldots, X_n$ to be known. Since this is not the case in practice, if $X_1, X_2, \ldots, X_n$ are not normally distributed, the limiting distribution of the statistics T_{n-1} can only be approximately normal (even if the variables are statistically independent). The quality of this approximation depends on the accuracy of the variance estimator.

APPENDIX B. SPECTRAL ANALYSIS OF VARIANCE

As was mentioned in Section 1, Heidelberger and Welch [1981a, 1981b] proposed estimating the variance of the estimator of mean in an autocorrelated sequence of observations from the smoothed averaged periodogram of this sequence; see Equations (45)–(50). The series of approximations they propose can be summarized as follows.

Procedure *SpectralVarianceAnalysis*;
{Preconditions:
$x_1, x_2, \ldots, x_{n_v}$ A sequence of n_v observations taken from a stationary process ($n_v \geq 100$, the default value: $n_v = 100$).
n_{ap} The number of points of the averaged periodogram used to fit it to a polynomial by applying the least squares procedure ($n_{ap} \leq n_v/4$, the default value $n_{ap} = 25$).
δ The degree of the polynomial fitted to the logarithm of the averaged periodogram (the default value $\delta = 2$).
C_σ A normalizing constant, chosen to make $\hat{p}_x(0)$ approximately unbiased (for the default values of n_{ap} and δ, take $C_\sigma = 0.882$ [Heidelberger and Welch 1981, Table 1], where these constants are denoted as K, d, and $C_1(K, d)$, respectively)}

Step 1
Calculate $2n_{ap}$ values of the periodogram of the sequence $x_1, x_2, \ldots, x_{n_v}$:

$$\Pi\left(\frac{j}{n_v}\right) = \left| \sum_{s=1}^{n_v} x_s \exp\left[-\frac{2\pi i(s-1)j}{n_v}\right] \right|^2 \Big/ n_v \quad \text{(B1)}$$

{*cf. Eqs. (45) and (46)*}
for $j = 1, 2, \ldots, 2n_{ap}$, and $i = \sqrt{-1}$;

Step 2
Calculate n_{ap} values of the function $\{L(f_j)\}$, for $j = 1, 2, \ldots, n_{ap}$; where $f_j = (4j-1)/2n_v$, and

$$L(f_j) = \log\{[\Pi((2j-1)/n_v) + \Pi(2j/n_v)]/2\}; \quad \text{(B2)}$$

Step 3
Apply the least squares extrapolation procedure {*see, for example, Press et al.* [1986, *p.* 509]} *for determining the coefficient* a_o *in the polynomial*

$$g(f) = \sum_{s=0}^{\delta} a_s f^s \qquad \text{(B3)}$$

fitted to the function $\{L(f_j) + 0.270\}$, $j = 1, 2, \ldots, n_{ap}$;
{*the value* a_o *is an unbiased estimate of log* $p_x(0)$}

Step 4
Calculate

$$\hat{p}_x(0) = C_o e^{a_o}; \qquad \text{(B4)}$$

$$\hat{\sigma}^2_{sp}[\bar{X}(n_v)] = \hat{p}_x(0)/n_v; \qquad \text{(B5)}$$

determine κ;
{*the degrees of freedom* κ *for the* χ^2 *distribution of* $\hat{\sigma}^2_{sp}[\bar{X}(n_v)]$, *for given* n_{ap} *and* δ, *are given in Heidelberger and Welch [1981, Table 1] and denoted there by* $C_2(K, d)$; *for* $n_v = 100$, $n_{ap} = 25$ *and* $\delta = 2$: $\kappa = 7$}
end *SpectralVarianceAnalysis*.

If an analyzed sequence of output data consists of n_v batch means $\bar{X}_1(m)$, $\bar{X}_2(m)$, $\ldots$, $X_{n_v}(m)$, each over m observations, instead of n_v individual observations x_1, x_2, $\ldots$, x_{n_v} as above, the batch means should replace individual observations in Equation (B1). Then the periodogram $\{\Pi(j/n)\}$ and the logarithm of the averaged periodogram $\{L(f_j)\}$ are functions of the sequence $\bar{X}_1(m)$, $\bar{X}_2(m)$, $\ldots$, $\bar{X}_{n_v}(m)$, and Equation (B4) gives us the estimate $\hat{p}_{\bar{X}(m)}(0)$, the spectral density function of this sequence of batch means at $f = 0$. Now, with n_v meaning the number of batch means, each over m observations, we obtain

$$\hat{p}_x(0) = m\hat{p}_{\bar{X}(m)}(0); \qquad \text{(B6)}$$

see Equation (49). Thus, the estimator of the variance of the mean $\bar{X}(n_v m)$, over $n_v m$ individual observations, is

$$\hat{\sigma}^2_{sp}[\bar{X}(n_v m)] = \frac{\hat{p}_{\bar{X}(m)}(0)}{n_v}. \qquad \text{(B7)}$$

The degrees of freedom can be determined as in the previous case; see the comment in Step 4 of the procedure.

ACKNOWLEDGMENTS

I wish to thank my colleagues Tim Bell, Rod Harries, and Don McNickle for helpful comments during preparation of this paper and Mehdi Asgarkhani and Tim Street for testing the sequential procedures presented. Special thanks are also due to two anonymous referees, whose comments greatly improved the final presentation.

REFERENCES

ABATE, J., AND WHITT, W. 1987a. Transient behavior of regulated brownian motion, I: Starting at the origin. *Adv. Appl. Probab. 19*, 560–598.

ABATE, J., AND WHITT, W. 1987b. Transient behavior of regulated brownian motion, II: Non-zero initial conditions. *Adv. Appl. Probab. 19*, 599–631.

ABATE, J., AND WHITT, W. 1987c. Transient behavior of the M/M/1 queue: Starting at the origin. *Queue. Syst. 2*, 41–65.

ABATE, J., AND WHITT, W. 1988. Transient behavior of the M/M/1 queue via Laplace transform. *Adv. Appl. Probab. 20*, 145–178.

ADAM, N. R. 1983. Achieving a confidence interval for parameters estimated by simulation. *Manage. Sci., 29*, 856–866.

AMEMIYA, T. 1973. Generalized least squares with an estimated autocovariance matrix. *Econometria 41*, 723–732.

ANDERSON, K. 1985. Some steady-state problems in simulation. In *Proceedings of the Annual Simulation Symposium* (Tampa, Fla.). IEEE Computer Society, New York, 79–93.

ANDREWS, R. W., AND SCHRIBER, T. J. 1978. Interactive analysis of output from GPSS-based simulations. In *Proceedings of the 1978 Winter Simulation Conference* (Miami Beach, Fla.). ACM/IEEE, New York, 267–278.

ANONUEVO, R., AND NELSON, B. L. 1986. Automated estimation and variance reduction for steady state simulation. In *Proceedings of the 1986 Winter Simulation Conference* (Washington, D.C.). ACM/IEEE, New York, 871–875.

Artificial intelligence and simulation: The diversity of applications. 1988. In *Proceedings of the SCS Multiconference* (San Diego, Calif.). Society for Computer Simulation.

ASGARKHANI, M., AND PAWLIKOWSKI, K. 1989. Simulation studies of mixed traffic on satellite channels using TDMA-reservation protocol. In *Proceedings of the 8th Phoenix International Conference on Computers and Communications* (Scottsdale, Ariz.). IEEE, Washington, D.C., 195–200.

BAIK, D., AND ZEIGLER, B. P. 1985. Performance evaluation of hierarchical distributed simulators. In *Proceedings of the 1985 Winter Simulation Conference* (San Francisco, Calif.). ACM/IEEE, New York, 421–427.

BALMER, D. W. 1987. Polishing the analysis of the statistical output of comparative simulation experiments. *Simulation, 49*, 123–126.

BANKS, J., AND CARSON, J. S. 1984. *Discrete Event System Simulation.* Prentice-Hall, Englewood Cliffs, N.J.

BARTLETT, M. S. 1946. On the theoretical specification of the sampling properties of autocorrelated time series. *J. Roy. Stat. Soc. Ser. B, 8*, 27–41.

BAUER, K. W., VENKATRAMAN, S., AND WILSON, J. R. 1987. Estimation procedures based on control variates with known covariance matrix. In *Proceedings of the 1987 Winter Simulation Conference* (Atlanta, Ga.). ACM/IEEE, New York, 334–341.

BEALL, C. W. 1982. A regression technique or determining steady state conditions in time series simulations. In *Proceedings of the 1982 Winter Simulation Conference* (San Diego, Calif.). ACM/IEEE, New York, 439–448.

BELL, P. C., AND O'KNEEFE, R. M. 1987. Visual interactive simulation: History, recent developments, and major issues. *Simulation, 49*, 109–116.

BHARATH-KUMAR, K., AND KERMANI, P. 1984. Performance evaluation tool (PET): An analysis tool for computer communication networks. *IEEE J. Select. Areas Commun. SAC-2*, 220–225.

BILLINGSLEY, P. 1968. *Convergence of Probability Measures.* Wiley, New York.

BLACKMAN, R. B., AND TUCKEY, J. W. 1958. The measurement of power spectra from the point of view of communications engineering (Part 1 and Part 2). *Bell Syst. Tech. J. 37*, 185–282, 485–569.

BLANC, J. P. C. 1985a. The transient behaviour of networks with infinite servers nodes. In *Performance '84*, E. Gelenbe, Ed. North Holland, Amsterdam, 159–174.

BLANC, J. P. C. 1985b. The relaxation time of two queueing systems in series. *Commun. Statist. Stochastic Models 1*, 1–16.

BLOMQVIST, N. 1967. The covariance function of the M/G/1 queueing system. *Skandinavisk Aktuarietidskrift, 6*, 157–174.

BLOMQVIST, N. 1970. On the transient behavior of the GI/G/1 waiting-times. *Skandinavisk Aktuarietidskrift, 53*, 118–129.

BLUM, A., ET AL. 1985. Experiments with decomposition of extended queueing network models. In *Modelling Techniques and Tools for Performance Analysis*, D. Potier, Ed. North-Holland, The Netherlands, 623–638.

BOX, G. E. P., AND JENKINS, G. M. 1970. *Time Series Analysis, Forecasting and Control.* Holden-Day, San Francisco.

BRATLEY, P., FOX, B. L., AND SCHRAGE, L. E. 1983. *A Guide to Simulation.* Springer-Verlag, New York.

BRILLINGER, D. R. 1973. Estimation of the mean of a stationary time series by sampling. *J. Appl. Prob., 10*, 419–431.

BRILLINGER, D. R. 1981. *Time Series: Data Analysis and Theory.* Holden-Day, San Francisco.

BROCKWELL, P. J., AND DAVIS, R. A. 1987. *Time Series: Theory and Methods.* Springer-Verlag, New York.

BULGREN, W. G. 1982. *Discrete System Simulation.* Prentice-Hall, Englewood Cliffs, N.J.

CALVIN, J. M. 1988. Covariance of regenerative mean and variance estimators for the Markov chains. In *Proceedings of the 1988 Winter Simulation Conference* (San Diego, Calif.). ACM/IEEE, New York, 473–475.

CARSON, J. S., AND LAW, A. M. 1980. Conservation Equations and Variance Reduction in Queueing Simulations. *Oper. Res., 28*, 535–546.

Catalog of simulation software. 1987. *Simulation, 49*, 165–181.

CHANDY, K. M., AND MISRA, J. 1981. Asynchronous distributed simulation via a sequence of parallel computations. *Commun. ACM, 24*, 198–205.

CHARNES, J. M., AND KELTON, W. D. 1988. A comparison of confidence region estimators for multivariate simulation output. In *Proceedings of the 1988 Winter Simulation Conference* (San Diego, Calif.). ACM/IEEE, New York, 458–465.

CHEN, D. R., AND SEILA, A. F. 1987. Multivariate inference in stationary simulation using batch means. In *Proceedings of the 1987 Winter Simulation Conference* (Atlanta, Ga.). ACM/IEEE, New York, 302–304.

CHENG, R. C. H. 1976. A note on the effect of initial conditions on a simulation run. *Oper. Res. Quart., 24*, 467–470.

CHENG, R. C. H. 1986. Variance reduction methods. In *Proceedings of the 1986 Winter Simulation Conference* (Washington, D.C.). ACM/IEEE, New York, 60–68.

CLARK, G. M. 1986. A Bonferroni selection procedure when using common numbers with unknown variances. In *Proceedings of the 1987 Winter Simulation Conference* (Washington, D.C.). ACM/IEEE, New York, 313–315.

COHEN, J. W. 1982. *The Single Server Queue.* North-Holland, Amsterdam.

CONWAY, R. W. 1963. Some tactical problems in digital simulation. *Manage. Sci., 10*, 47–61.

CONWAY, R. W., JOHNSON, B. M., AND MAXWELL, W. C. 1959. Some problems of digital systems simulation. *Manage. Sci. 6*, 92–110.

COTTRELL, M., FORT, J. C., AND MALGOUYRES, G. 1983. Large deviations and rare events in the study of stochastic algorithms. *IEEE Trans. Automatic Control, AC-28*, 907–918.

COX, D. R., AND SMITH, S. W. L. 1961. *Queues.* Wiley, New York.

CRANE, M., AND IGLEHART, D. L. 1974. Simulating stable stochastic systems: I. General multiserver queues. *J. ACM, 21*, 103–113.

CRANE, M. A., AND IGLEHART, D. L. 1975a. Simulating stable stochastic systems: III. Regenerative processes and discrete-event simulations. *Oper. Res., 23*, 33–45.

CRANE, M. A., AND IGLEHART, D. L. 1975b. Simulating stable stochastic systems. IV: Approximation techniques. *Manage. Sci., 21*, 1215–1224.

CRANE, M. A., AND LEMOINE, A. J. 1977. *An Introduction to the Regenerative Method for Simulation Analysis.* Lecture Notes in Control and Information Science, No. 4, A. W. Balakrishnan and M. Thoma, Eds. Springer-Verlag, New York.

DALEY, D. J. 1968. The serial correlation coefficients of waiting times in a stationary simple server queue. *J. Austral. Math. Soc., 8,* 683–699.

DAMERDJI, H. 1987. On strong consistency of the variance estimator. In *Proceedings of the 1987 Winter Simulation Conference* (Atlanta, Ga.). ACM/IEEE, New York, 305–308.

DECEGAMA, A. 1987. Parallel processing simulation of large computer networks. In *Proceedings of the 1987 Symposium on the Simulation of Computer Networks* (Colorado). ACM/IEEE, New York, 51–62.

DONNELLY, J. N., AND SHANNON, R. E. 1981. Minimum mean-squared error estimates for simulation experiments. *Commun. ACM, 24,* 253–259.

DUERSCH, R. R., AND SCHRUBEN, L. W. 1986. An interactive run length control for simulations on PCs. In *Proceedings of the 1986 Winter Simulation Conference* (Washington, D.C.). ACM/IEEE, New York, 866–870.

DUKET, S. D., AND PRITSKER, A. A. B. 1978. Examination of simulation output using spectral methods. *Math. Comput. Simul., 20,* 53–60.

EMSHOFF, J. R., AND SISSON, R. L. 1970. *Design and Use of Computer Simulation Models.* Macmillan, New York.

FISHMAN, G. S. 1967. Problems in the statistical analysis of simulation experiments: The comparison of means and the length of sample records. *Commun. ACM, 10,* 94–99.

FISHMAN, G. S. 1971. Estimating sample size in computing simulation experiments. *Manage. Sci., 18,* 21–38.

FISHMAN, G. S. 1972. Bias considerations in simulation experiments. *Oper. Res., 20,* 685–790.

FISHMAN, G. S. 1973a. *Concepts and Methods in Discrete Event Digital Simulation.* John Wiley, New York.

FISHMAN, G. S. 1973b. Statistical analysis for queueing simulation. *Manage. Sci., 20,* 363–369.

FISHMAN, G. S. 1974. Estimation in multiserver queueing simulations. *Oper. Res., 22,* 72–78.

FISHMAN, G. S. 1977. Achieving specific accuracy in simulation output analysis. *Commun. ACM, 20,* 310–315.

FISHMAN, G. S. 1978a. *Principles of Discrete Event Simulation.* John Wiley, New York.

FISHMAN, G. S. 1978b. Grouping observations in digital simulation. *Manage. Sci., 24,* 510–521.

FISHMAN, G. S., AND KIVIAT, P. J. 1967. The analysis of simulation-generated time series. *Manage. Sci., 13,* 525–557.

FOLEY, R. D., AND GOLDSMAN, D. 1988. Confidence intervals and orthonormally weighted standardized time series. In *Proceedings of the 1988 Winter Simulation Conference* (San Diego, Calif.). ACM/IEEE, New York, 422–424.

FOX, B. 1978. Estimation and simulation. *Manage. Sci., 24,* 860–861.

FRIEDMAN, L. W., AND FRIEDMAN, H. H. 1986. Comparing simulated alternatives using a distribution-free statistic with blocking by random number stream. *Simulation, 48,* 68–70.

FRIEDMAN, L. W. 1984. Multivariate simulation output analysis: Past, present, and future. In *Proceedings of the 1984 Winter Simulation Conference* (Dallas, Tex.). ACM/IEEE, New York, 277–281.

FROST, V. S., LAURE, W. W., AND SHANMUGAN, K. S. 1988. Efficient techniques for simulation of computer communication networks. *IEEE J. Selected Areas Commun., SAC-8,* 146–157.

FUJIMOTO, R. M. 1988. Performance measurements of distributed simulation strategies. In *Proceedings of Distributed Simulation 1988* (San Diego, Calif.). Society for Computer Simulation, 14–20.

GAFARIAN, A. V., ANCKER C. J., AND MORISAKU, T. 1978. Evaluation of commonly used rules for detecting "steady state" in computer simulation. *Naval Res. Logist. Quart. 78,* 511–529.

GATES, B., CAMMARATA, S., AND ROTHENBERG, J. 1988. A demon facility for object-oriented simulation language. In *Proceedings of the 1988 Summer Simulation Conference* (Seattle, Wash.). Society for Computer Simulation, San Diego, Calif., 667–673.

GEISLER, M. A. 1964. The size of simulation samples required to compute certain inventory characteristics with stated precision and confidence. *Manage. Sci., 10,* 261–271.

GLYNN, P. W. 1982. Coverage error for confidence intervals arising in simulation output analysis. In *Proceedings of the 1982 Winter Simulation Conference* (San Diego, Calif.). ACM/IEEE, New York, 369–375.

GLYNN, P. W., AND IGLEHART, D. L. 1985. Large-sample theory for standardized time series: An overview. In *Proceedings of the 1985 Winter Simulation Conference* (San Francisco, Calif.). ACM/IEEE, New York, 129–134.

GLYNN, P. W., AND WHITT, W. 1989. Indirect estimation via $L = \lambda W$. *Manage. Sci., 36,* 82–103.

GOLDSMAN, D. 1983. Ranking and selection in simulation. In *Proceedings of the 1983 Winter Simulation Conference* (Arlington, Va.). ACM/IEEE, New York, 387–393.

GOLDSMAN, D. 1986. Tutorial on indifference-zone normal means ranking and selection procedures. In *Proceedings of the 1986 Winter Simulation Conference* (Washington, D.C.). ACM/IEEE, New York, 370–375.

GOLDSMAN, D., AND MEKETON, M. S. 1985. A comparison between standardized time series and overlapping batched means. In *Proceedings of the*

1985 Winter Simulation Conference. IEEE, New York, 208–210.

GOLDSMAN, D., AND MEKETON, M. S. 1990. A comparison of several variance estimators. *Oper. Res.* (to be published).

GOLDSMAN, D., KANG, K., AND SARGENT, R. G. 1986. Large and small sample comparisons of various variance estimators. In *Proceedings of the 1986 Winter Simulation Conference* (Washington, D.C.). ACM/IEEE, New York, 278–284.

GOLDSMAN, D., AND SCHRUBEN, L. 1984. Asymptotic properties of some confidence interval estimators for simulation output. *Manage. Sci., 30,* 1217–1225.

GORDON, G. 1969. *System Simulation.* Prentice Hall, Englewood Cliffs, N.J.

GROSS, D., AND HARRIS, C. M. 1974. *Fundamentals of Queueing Theory.* John Wiley, New York.

GUNTHER, F. L., AND WOLFF, R. W. 1980. The almost regenerative method for stochastic system simulations. *Oper. Res., 28,* 375–386.

HADDOCK, J. 1987. An expert system framework based on a simulation generator. *Simulation, 49,* 45–53.

HAHN, G. J. 1985. More intelligent statistical software and statistical expert systems: Future directions. *Am. Stat., 39,* 1–16.

HAMMING, R. W. 1962. *Numerical Methods for Scientists and Engineers.* McGraw-Hill, New York.

HAND, D. J. 1985. Statistical expert systems: Necessary attributes. *J. Appl. Stat., 12,* 19–27.

HANNAN, E. J. 1970. *Multiple Time Series.* John Wiley, New York.

HARLING, J. 1985. Simulation techniques in operations research: A review. *Oper. Res., 33,* 307–319.

HEIDELBERGER, P. 1979. A variance reduction technique that increases regeneration frequency. In *Current Issues in Computer Simulation.* Academic Press, New York, 257–269.

HEIDELBERGER, P. 1986. Statistical analysis of parallel simulation. In *Proceedings of the 1986 Winter Simulation Conference* (Washington, D.C.). ACM/IEEE, New York, 290–295.

HEIDELBERGER, P., AND LEWIS, P. A. W. 1981. Regression: Adjusted estimates for regenerative simulations, with graphics. *Commun. ACM, 24,* 260–273.

HEIDELBERGER, P., AND LEWIS, P. A. W. 1984. Quantile estimation in dependent sequences. *Oper. Res., 32,* 185–209.

HEIDELBERGER, P., AND WELCH, P. D. 1981a. A spectral method for confidence interval generation and run length control in simulations. *Commun. ACM, 24,* 233–245.

HEIDELBERGER, P., AND WELCH, P. D. 1981b. Adaptive spectral methods for simulation output analysis. *IBM J. Res. Dev., 25,* 860–876.

HEIDELBERGER, P., AND WELCH, P. D. 1983. Simulation run length control in the presence of an initial transient. *Oper. Res., 31,* 1109–1144.

HO, Y. C. 1987. Performance evaluation and perturbation analysis of discrete event dynamic systems. *IEEE Trans. Automat. Control, AC-32,* 563–572.

HO, Y. C., AND CAO, X. 1983. Perturbation analysis and optimization of queueing networks. *J. Optim. Theory Appl., 40,* 559–582.

HO, Y. C., CAO, X., AND CASSANDRAS, C. 1983. Infinitesimal and finite perturbation analysis for queueing networks. *Automatica, 4,* 439–445.

IGLEHART, D. L. 1975. Simulating stable stochastic systems. V: Comparison of ratio estimators. *Naval Res. Logist. Quart., 22,* 553–565.

IGLEHART, D. L. 1976. Simulating stable stochastic systems VI: Quantile estimation. *J. ACM* 347–360.

IGLEHART, D. L. 1978. The regenerative method for simulation analysis. In *Current Trends in Programming Methodology.* Vol. III, *Software Modeling,* K. M. Chandy and P. T. Yeh, Eds., Prentice Hall, Englewood Cliffs, N.J., 52–71.

IZYDORCZYK, J., MIERZWA, J., AND WOLISZ, A. 1984. A control variable method for increasing the efficiency of multiple access protocol simulation. In *Performance of Computer-Communication Systems,* H. Rudin and W. Bux, Eds., North-Holland, Amsterdam, 95–108.

JACKMAN, J., AND MEDEIROS, D. J. 1988. A graphical methodology for simulating communication networks. *IEEE Trans. Commun., ACOM-36,* 459–464.

JAIN, R., AND CHLAMTAC, I. 1985. The P^2 algorithm for dynamic calculation of quantiles and histograms without storing observations. *Commun. ACM, 28,* 1076–1085.

JENKINS, G. M., AND WATTS, D. G. 1968. *Spectral Analysis and Its Applications.* Holden-Day, San Francisco.

KANG, K., AND GOLDSMAN, D. 1985. The correlation between mean and variance estimators. In *Proceedings of the 1985 Winter Simulation Conference* (San Francisco, Calif.). ACM/IEEE, New York, 211–216.

KELTON, W. D. 1986. Statistical design and analysis. In *Proceedings of the 1986 Winter Simulation Conference* (Washington, D.C.). ACM/IEEE, New York, 45–51.

KELTON, W. D. 1985. Transient exponential-Erlang queues and steady-state simulation. *Commun. ACM, 28,* 741–749.

KELTON, W. D., AND LAW. A. M. 1985. The transient behaviour of the M/M/S queue, with implications for steady-state simulation. *Oper. Res., 33,* 378–396.

KELTON, W. D., AND LAW, A. M. 1984. An analytical evaluation of alternative strategies in steady-state simulation. *Oper. Res., 32,* 169–184.

KELTON, W. D., AND LAW, A. M. 1983. A new approach for dealing with the startup problem in discrete event simulation. *Naval Res. Logist. Quart., 30,* 641–658.

KERCKHOFFS, E. J. H., AND VANSTEENKISTE, G. C. 1986. The impact of advanced information processing on simulation: An illustrative review. *Simulation, 48*, 17–26.

KIEFER, J., AND WOLFOWITZ, J. 1955. On the theory of queues with many servers. *Trans. Am. Math. Soc., 78*, 1–18.

KLEIJNEN, J. P. C. 1974. *Statistical Techniques in Simulation*, Vol. 1. Marcel Dekker, New York.

KLEIJNEN, J. P. C. 1979. The role of statistical methodology in simulation. In *Methodology in Systems Modelling and Simulation*, B. P. Zeigler, M. S. Elzas, G. J. Klir, and T. I. Oren, Eds. North-Holland, Amsterdam.

KLEIJNEN, J. P. C. 1987. *Statistical Tools for Simulation Practitioners*. Marcel Dekker, New York.

KLEIJNEN, J. P. C., VAN DER VEN, R., AND SAUNDER, B. 1982. Testing independence of simulation subruns: A note on the power of the von Neumann test. *Eur. J. Oper. Res., 9*, 92–93.

KLEINROCK, L. 1976. *Queueing Systems*. Vol. 2, *Computer Applications*. John Wiley, New York.

KNAPP, V. 1986. The smalltalk simulation environment. In *Proceedings of the 1986 Winter Simulation Conference* (Washington, D.C.). ACM/IEEE, New York, 125–128.

KREUTZER, W. 1986. *System Simulation. Programming Styles and Languages*. Addison Wesley, Sydney.

KRISHNAMURTHY, M., CHANDRASEKARAN, U., AND SHEPPARD, S. 1985. Two approaches to the implementation of a distributed simulation system. In *Proceedings of the 1985 Winter Simulation Conference* (San Francisco, Calif.). ACM/IEEE, New York, 435–443.

KUMAR, D. 1986. A novel approach to sequential simulation. *IEEE Softw., 3*, 25–33.

KUROSE, J. F., AND MOUFTAH, H. T. 1988. Computer-aided modeling, analysis, and design of communication networks. *IEEE J. Select. Areas Commun., SAC-6*, 130–145.

LAVENBERG, S. S., AND SAUER, C. H. 1977. Sequential stopping rules for the regenerative method of simulation. *IBM J. Res. Dev., 21*, 545–558.

LAVENBERG, S. S., AND WELCH, P. D. 1981. A perspective on the use of control variables to increase the efficiency of Monte Carlo simulations. *Manage. Sci., 27*, 322–335.

LAVENBERG, S. S., MOELLER, T. L., AND WELCH, P. D. 1982. Statistical results on control variables with application to queueing network simulation. *Oper. Res., 30*, 182–202.

LAVENBERG, S. S., ET AL. 1981. The initial transient in steady state simulation (panel discussion). In *Proceedings of the 1981 Winter Simulation Conference* (Atlanta, Ga.). ACM/IEEE, New York, 113–120.

LAW, A. M. 1975. Efficient estimators for simulated queueing systems. *Manage. Sci., 22*, 30–41.

LAW, A. M. 1977. Confidence interval in discrete event simulation: A comparison of replication and batch means. *Naval Res. Logist. Quart., 24*, 667–678.

LAW, A. M. 1980. Statistical analysis of the output data from terminating simulations. *Naval Res. Logist. Quart., 27*, 131–143.

LAW, A. M. 1983. Statistical analysis of simulation output data. *Oper. Res., 31*, 983–1029.

LAW, A. M., AND CARSON, J. C. 1979. A sequential procedure for determining the length of a steady state simulation. *Oper. Res., 27*, 1011–1025.

LAW, A. M., AND KELTON, W. D. 1982a. *Simulation, Modeling and Analysis*. McGraw-Hill, New York.

LAW, A. M., AND KELTON, W. D. 1982b. Confidence intervals for steady state simulations. II: A survey of sequential procedures. *Manage. Sci., 28*, 550–562.

LAW, A. M., AND KELTON, W. D. 1984. Confidence intervals for steady state simulations. I: A survey of fixed sample size procedures. *Oper. Res.* 1221–1239.

MACNAIR, E. A. 1985. An introduction to the research queueing package. In *Proceedings of the 1985 Winter Simulation Conference* (San Francisco, Calif.). ACM/IEEE, New York, 257–262.

MADANSKY, A. 1976. Optimal initial conditions for a simulation problem. *Oper. Res., 24*, 172–577.

MARKS, N. B. 1981. Further investigation into spectral analysis for confidence intervals in steady state simulation. In *Proceedings of the 1981 Winter Simulation Conference* (Atlanta, Ga.). ACM/IEEE, New York, 461–464.

MEKETON, M. S. 1980. The variance time-curve: Theory, estimation, and application. Ph.D. dissertation, School of Operations Research and Industrial Engineering, Cornell Univ., Ithaca, N.Y.

MEKETON, M. S., AND HEIDELBERGER, P. 1982. A renewal theoretic approach to bias reduction on regenerative simulations. *Manage. Sci., 28*, 173–181.

MEKETON, M. S., AND SCHMEISER, B. 1984. Overlapping batch means: Something for nothing? In *Proceedings of the 1984 Winter Simulation Conference* (Dallas, Tex.). ACM/IEEE, New York, 227–230.

MELLICHAMP, J. M., AND PARK, Y. H. 1989. A statistical expert system for simulation analysis. *Simulation, 51*, 134–139.

MILLER, R. G. 1974. The Jackknife: A Review. *Biometrika, 61*, 1–15.

MILLS, R. 1987. *Statistical Analysis of Steady State Simulation Output Data with SIMSCRIPT II.5*. CACI, Los Angeles.

MINH, D. L. 1987. Simulating GI/G/k queues in heavy traffic. *Manage. Sci., 34*, 1192–1199.

MISRA, J. 1986. Distributed discrete event simulation. *ACM Comput. Surv., 18*, 39–66.

MITRANI, I. 1982. Simulation techniques for discrete event systems. Cambridge University Press, Cambridge, U.K.

MORSE, P. M. 1955. Stochastic processes of waiting lines. *Oper. Res., 1*, 255–261.

MURRAY, J. R., AND KELTON, W. D. 1988. Initializing for bias reduction: Some analytical considerations. In *Proceedings of the 1988 Winter Simulation Conference* (San Diego, Calif.). ACM/IEEE, New York, 546–548.

NELSON, B. L., AND SCHMEISER, B. W. 1986. Decomposition of some well-known variance reduction techniques. *J. Stat. Comput. Simul., 23*, 183–209.

NEWELL, G. F. 1971. *Applications of Queueing Theory.* Chapman & Hall, London.

NICOL, D. M. 1988. Parallel discrete event simulation of FCFS stochastic queueing networks. In *Parallel Programming: Experience with Applications, Languages and Systems*, ACM SIGPLAN, 124–137.

NICOL, D. M. 1988. High performance parallelized discrete-event simulation of stochastic queueing networks. In *Proceedings of the 1988 Winter Simulation Conference* (San Diego, Calif.). ACM/IEEE, New York, 237–241.

NIELSEN, N. R. 1986. Expert systems and simulation. In *Computer Networks and Simulation III*, S. Schoemaker, Ed., North-Holland, Amsterdam, 41–52.

ODONI, A. R., AND ROTH, E. 1983. An empirical investigation of the transient behaviour of stationary queueing systems. *Oper. Res., 34*, 306–314.

O'KEEFE, R. M. 1986. Simulation and expert systems: A taxonomy and some examples. *Simulation, 48*, 10–16.

ÖREN, T. I., AND ZEIGLER, B. P. 1987. Artificial intelligence in modelling and simulation: Directions to explore. *Simulation, 49*, 131–134.

PAREKH, S., AND WALRAND, J. 1989. A quick simulation of excessive backlogs in networks of queues. *IEEE Trans. Automatic Control, AC-34*, 54–66.

PARK, S. K., AND MILLER, K. W. 1988. Random number generators: Good ones are hard to find. *Commun. ACM, 31*, 1192–1201.

PARZEN, E. 1961. Mathematical consideration in the estimation of spectra. *Technometrics, 3*, 167–190.

PAWLIKOWSKI, K., AND ASGARKHANI, M. 1988. Sequential procedures in simulation studies of satellite protocols. In *Proceedings of the ITC'12* (Torino, Italy), Int. Advisory Council of ITC vol. 6, 4.3B.3.1–4.3B.3.7. (Also in *Teletraffic Science for New Cost-Effective Systems, Systems and Services.* Vol. 12, *Studies in Telecommunication.* North-Holland, Amsterdam, 1989.)

PAYNE, J. A. 1982. *Introduction to Simulation.* McGraw-Hill, New York.

PIDD, M. 1984. *Computer Simulation in Management Science.* Wiley, Chichester.

POTIER, D. 1984. New user's introduction to QNAP2. Rapport Technique INRIA, No. 40.

POTIER, D. 1986. The modeling package QNAP2 and applications to computer networks simulation. In *Computer Networks and Simulation III.* North-Holland, Amsterdam, 235–265.

PRESS, W. H., FLANNERY, B. P., TENKOLSKY, S. A. AND VATTERLING, W. T. 1986. *Numerical Recipes. The Art of Scientific Computing.* Cambridge University Press, Cambridge, U.K.

RAATIKAINEN, K. E. E. 1987. Simultaneous estimation of several percentiles. *Simulation, 49*, 159–164.

RAATIKAINEN, K. E. E. 1988. Validating percentiles of response times in queueing network models. In *Proceedings of 1988 Summer Simulation Conference* (Seattle, Wash.). Society for Computer Simulation, San Diego, Calif., 136–141.

REDDY, R. 1987. Epistemology of knowledge based simulation. *Simulation, 48*, 162–166.

REIMAN, M. I., AND WEISS, A. 1986. Sensitivity analysis via likelihood ratios. In *Proceedings of the 1986 Winter Simulation Conference* (Washington, D.C.). ACM/IEEE, New York, 285–289.

REYNOLDS, P. F. 1988. A spectrum of options for parallel simulation. In *Proceedings of the 1988 Winter Simulation Conference* (San Diego, Calif.). ACM/IEEE, New York, 325–332.

ROTH, E. 1985. A relaxation time heuristic for reducing initialization bias in simulation of $M/E_k/1$ queueing systems. In *Proceedings of the 1985 Winter Simulation Conference* (San Francisco, Calif.). ACM/IEEE, New York, 263–268.

ROTH, E., AND RUTAN, A. H. 1985. A relaxation time approach for reducing initialization bias in simulation. In *Proceedings, Annual Simulation Symp.* (Tampa, Fla.). IEEE Computer Society, New York, 189–203.

RUBINSTEIN, R. Y. 1989. Sensitivity analysis and performance extrapolation for computer simulation models. *Oper. Res., 37*, 72–81.

RUBINSTEIN, R. Y., AND MARCUS, R. 1985. Efficiency of multivariate control variates in Monte Carlo simulation. *Oper. Res., 33*, 661–677.

RUIZ-MIER, S., AND TALAVAGE, J. 1987. A hybrid paradigm for modeling of complex systems. *Simulation, 49*, 135–141.

SARGENT, R. G. 1986. The use of graphical models in model validation. In *Proceedings of the 1986 Winter Simulation Conference* (Washington, D.C.). ACM/IEEE, New York, 237–241.

SAUER, C. H., MACNAIR, E. A., AND KUROSE, J. K. 1984. Queueing network simulations of computer communications. *IEEE J. Select. Areas Commun., SAC-1*, 203–220.

SCHMEISER, B. 1982. Batch size effects in the analysis of simulation output. *Oper. Res., 30*, 556–568.

SCHMEISER, B., AND SONG, W. T. 1987. Correlation among estimators of the variance of the sample mean. In *Proceedings of the 1987 Winter Simulation Conference* (Atlanta, Ga.). ACM/IEEE, New York, 309–317.

SCHMIDT, J. W., AND HO, C. 1988. An algorithm for testing dependence of simulation output data. In *Proceedings of the 1988 Winter Simulation Conference* (San Diego, Calif.). ACM/IEEE, New York, 532–539.

SCHRIBER, T. 1974. *Simulation Using GPSS.* Wiley, New York.

SCHRIBER, T. J., AND ANDREWS, R. W. 1979. Interactive analysis of simulation output by the method of batch means. In *Proceedings of the 1979 Winter Simulation Conference* (San Diego, Calif.). ACM/IEEE, New York, 512–524.

SCHRIBER, T. J., AND ANDREWS, R. W. 1981. A conceptual framework for research in the analysis of simulation output. *Commun. ACM, 24,* 218–232.

SCHRUBEN, L. W. 1980. A coverage function for interval estimators of simulation response. *Manage. Sci., 26,* 18–27.

SCHRUBEN, L. W. 1981. Control of initialization bias in multivariate simulation response. *Commun. ACM, 24,* 246–252.

SCHRUBEN, L. W. 1982. Detecting initialisation bias in simulation output. *Oper. Res.* 569–590.

SCHRUBEN, L. W. 1983. Confidence interval estimation using standardised time series. *Oper. Res., 30,* 1090–1108.

SCHRUBEN, L. W. 1985. Overview of standardized time series. In *Proceedings of the 1985 Winter Simulation Conference* (San Francisco, Calif.). ACM/IEEE, New York, 115–118.

SCHRUBEN, L. W. 1987. Using simulation to solve problems: A tutorial on the analysis of simulation output. In *Proceedings of the 1987 Winter Simulation Conference* (Atlanta, Ga.). ACM/IEEE, New York, 40–42.

SCHRUBEN, L. W., SINGH, H., AND TIERNEY, L. 1983. Optimal tests for initialisation bias in simulation output. *Oper. Res., 31,* 1167–1178.

SEILA, A. F. 1982. A batching approach to quantile estimation in regenerative simulation. *Manage. Sci., 28,* 573–581.

SEILA, A. F. 1982. Multivariate estimation in regenerative simulation. *Oper. Res. Lett., 1,* 153–156.

SEILA, A. F. 1983. Multivariate estimation in simulation. In *Proceedings of the 1983 Winter Simulation Conference* (Arlington, Va.). ACM/IEEE, New York, 693–696.

SHAHABUDDIN, P., ET AL. 1988. Variance reduction in mean time to failure simulation. In *Proceedings of the 1988 Winter Simulation Conference* (San Diego, Calif.). ACM/IEEE, New York, 491–499.

SHANNON, R. E. 1981. Test for the verification and validation of computer simulation models. In *Proceedings of the 1981 Winter Simulation Conference* (Atlanta, Ga.). ACM/IEEE, New York, 573–577.

SHANTHIKUMAR, J., AND SARGENT, R. 1983. A unifying view of hybrid simulation/analytic models and modeling. *Oper. Res., 31,* 1030–1052.

SHAPIRO, S. S., AND WILK, M. B. 1965. An analysis of variance test for normality (complete) samples. *Biometrika, 52,* 591–611.

SHEDLER, G. S. 1987. *Regeneration and Networks of Queues.* Springer-Verlag, New York.

SOLOMON, S. L. 1983. *Simulation of Waiting-Line Systems.* Prentice-Hall, Englewood Cliffs, N.J.

SONG, W. T. 1988. Estimators of the Variance of the Simple Mean: Quadratic Forms, Optimal Batch Sizes, and Linear Combinations. Ph.D. Dissertation, School of Industrial Engineering, Purdue University, West Lafayette, Indiana, 1988.

SONG, W. T., AND SCHMEISER, B. 1988. Minimal MSE linear combinations of variance estimators of the sample mean. In *Proceedings of the 1988 Winter Simulation Conference* (San Diego, Calif.). ACM/IEEE, New York, 414–421.

STAIRMAND, M. C., AND KREUTZER, W. 1988. POSE: A process-oriented simulation environment embedded in scheme. *Simulation, 50,* 143–153.

SURI, R., AND ZAZANIS, M. A. 1988. Perturbation analysis gives strongly consistent sensitivity estimates for the M/G/1 queue. *Manage. Sci., 35,* 39–64.

TOCHER, K. D. 1963. *The Art of Simulation.* The English Universities Press Ltd., London.

TRIVEDI, K. S. 1982. *Probability and Statistics with Reliability, Queueing, and Computer Science Applications.* Prentice-Hall, Englewood Cliffs, N.J.

TURNQUIST, M. A., AND SUSSMAN, J. M. 1977. Toward guidelines for designing experiments in queuing simulation. *Simulation, 28,* 137–144.

UNGER, B. W. 1988. Distributed simulation. In *Proceedings of the 1988 Winter Simulation Conference* (San Diego, Calif.). ACM/IEEE, New York, 198–205.

VASSILACOPOULOS, G. 1989. Testing for initialization bias in simulation output. *Simulation, 52,* 151–153.

VELAYAS, J. M., AND LEVARY, R. R. 1987. Validation of simulation models using decision theory. *Simulation, 48,* 87–92.

VENKATRAMAN, S., AND WILSON, J. R. 1985. Using path control variates in activity network simulation. In *Proceedings of the 1985 Winter Simulation Conference* (San Francisco, Calif.). ACM/IEEE, New York, 217–222.

VENKATRAMAN, S., AND WILSON, J. R. 1986. The efficiency of control variates in multiresponse simulation. *Oper. Res. Lett. 5,* 37–42.

VON NEUMANN, J. 1941. Distribution of the ratio of the mean square successive difference to the variance. *Ann. Math. Stat., 12,* 367–395.

WAGNER, D. B., AND LAZOWSKA, E. D. 1989. Parallel simulation of queueing networks: Limitations and potentials. *Perform. Eval. Rev., 17,* 146–155.

WAHBA, G. 1980. Automatic smoothing of the log periodogram. *J. Am. Stat. Assoc., 57,* 122–132.

WALRAND, J. 1988. *Computer Performance and Reliability: Proceedings of the 2nd Int. NCPR Workshop* (Rome). North-Holland, Amsterdam, 275–286.

WELCH, P. D. 1983. Statistical Analysis of Simulation Results. In *Computer Performance Modelling Handbook*, S. S. Lavenberg, Ed. Academic Press, New York, Chap. 6.

WELCH, P. D. 1987. On the relationship between batch means, overlapping batch means and spectral estimation. In *Proceedings of the 1987 Winter Simulation Conference* (Atlanta, Ga.). ACM/IEEE, New York, 320–323.

WILSON, J. R. 1984. Variance reduction techniques for digital simulation. *Am. J. Math. Manage. Sci., 4*, 277–312.

WILSON, J. R. 1983. Variance reduction: The current state. *Math. Comput. Simul., 26*, 55–59.

WILSON, J. R., AND PRITSKER, A. A. B. 1978a. A survey of research on the simulation startup problem. *Simulation, 31*, 55–58.

WILSON, J. R., AND PRITSKER, A. A. B. 1978b. Evaluation of startup policies in simulation experiments. *Simulation, 31*, 79–89.

YANG, W., AND NELSON, B. L. 1988. Multivariate estimation and variance reduction in terminating and steady state simulation. In *Proceedings of the 1988 Winter Simulation Conference* (San Diego, Calif.). ACM/IEEE, New York, 466–472.

ZEIGLER, B. P. 1987. Hierarchical, modular discrete-event modelling in an object-oriented environment. *Simulation, 49*, 219–230.

Received July 1988; final revision accepted December 1989.